DOCUMENT RESUME

ED 040 591 56 EM 008 175

AUTHOR	Karis, Charles; And Others
TITLE	The Interactive Effect of Responses per Frame, Response Mode, and Response Confirmation on Intraframe S-R Association Strength. Final Report.
INSTITUTION	Northeastern Univ., Boston, Mass.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-5-0773
PUB DATE	May 70
GRANT	OEG-7-31-0570-280
NOTE	85p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$4.35
DESCRIPTORS	Medical Education, *Program Content, *Programed Instruction, *Response Mode

ABSTRACT

To investigate the role of response mode, confirmation procedure, and frame content variables in linear self-instructional programs, a 384-frame medical terminology program was developed, tested, and validated. The program was then administered during periods of four consecutive days to 450 freshman students who were randomly assigned to a group using one of 18 versions of the program. Criterion measures consisted of four daily unit tests, a post-program comprehensive test, and a delayed retention test; items in each test required either the reproduction of medical terms or the definition of medical terms. Findings from both daily and comprehensive tests showed that overt response was superior to covert response and that reading responses was definitely a function of the reproduction accuracy required on criterion items, for reproduction of medical terms. Also, while variations in frame size resulted in no significant difference among the groups for these same two tests, the variation of number of responses required per frame produced a significant effect in favor of multiple responses when test items required medical term responses. (Author/SP)

U.S. DEPARTMENT OF HEALTH EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED040591

FINAL REPORT
Project No. 5-0773
Grant No. OE-7-31-0570-280

THE INTERACTIVE EFFECT OF RESPONSES PER FRAME, RESPONSE MODE, AND RESPONSE CONFIRMATION ON INTRAFRAME S-R ASSOCIATION STRENGTH

Charles Karis
Northeastern University
360 Huntington Avenue
Boston, Massachusetts 02115

May 1970

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EM 008 175

ED040591

Final Report
Project No. 5-0773
Grant No. OE-7-31-0570-280

The Interactive Effect of Responses Per Frame, Response Mode, and Response Confirmation on Intraframe S-R Association Strength

Charles Karis, Alvin Kent, and James E. Gilbert

Northeastern University

Boston, Massachusetts 02115

May 1970

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

ABSTRACT

Previous studies investigating the role of response mode, confirmation procedure, and frame content variables in linear self-instructional programs have left a number of important questions unresolved. Under what conditions does the written response have a differential effect on achievement? Is there any need for a confirmation procedure in programmed instruction? Should any restrictions be placed upon the amount of informational content in program frames? Is there any evidence indicating a need for limiting the number of responses required by each program frame? Are these considerations involved in any important interactive effects?

The present investigation was organized into three separate but related studies. Studies 1 and 2, concerned with response mode and confirmation procedures respectively, provided information for refining the experimental design for Study 3. The final study investigated the main and interactive effects of response mode, confirmation, frame content and number of responses per frame on program performance and retention test achievement.

In Study 1, a 197-frame program on medical terminology was developed and administered to 50 unpaid volunteer undergraduate students. Achievement was measured by a 120-item test given seven days after program completion. A scoring procedure, developed for analyzing response reproduction accuracy, was used to evaluate the test results. It was found that subjects assigned to an overt response group reproduced medical terms with significantly greater proficiency than those assigned to a covert response group, but only when the scoring criterion required errorless spelling. Removing this requirement resulted in nonsignificant differences between the groups. Nonsignificant differences between the response mode groups were also found on test items which required definitions of medical terms as responses.

As expanded version of the same program on medical terminology, now consisting of 378 frames, was administered to 96 paid freshman engineering students volunteering for the second study. Subjects were randomly assigned to either one of four groups in which the format of the confirmation procedure was varied, or to a fifth group in which confirmation was omitted altogether. Programmed learning sessions were scheduled for distribution over four consecutive days. Achievement was measured by four daily unit tests, and a comprehensive post-program test administered on the fifth consecutive day. Each test was composed of items requiring the written recall of the medical terms and definitions presented in the program. Additional items requiring definitions of medical terms not encountered in the program, were included as part of the comprehensive test to measure proficiency in the application of the medical word-building principles taught by the program. No significant differences were found for any of the groups on any of the tests administered. Varying confirmation procedures or withholding confirmation altogether appeared to have no effect on either the recall or the reproduction accuracy of medical terms or their definitions. In addition, an analysis of time to complete the program and the number of program

errors associated with the assigned confirmation procedure indicated no significant differences among the five groups.

In Study 3, the medical terminology program (384 frames), now completely validated and revised through analysis of error data collected during the first two studies, was administered to 450 paid volunteer freshman students during periods of four consecutive days. Subjects were randomly assigned to a group using one of 18 versions of the program. Sixteen versions were used to compare the main and interactive treatment effects of overt vs. covert responding, confirmation vs. non-confirmation, limited frame content vs. expanded frame content, and single vs. multiple frame responses in a 2 x 2 x 2 x 2 factorial design. Two additional versions, representing reading programs, were included to provide control data. Criterion measures consisted of four daily unit tests, a post-program comprehensive test, and a delayed retention test administered approximately one month later. Items in each test required either the reproduction of medical terms or the definition of medical terms. For both the daily and comprehensive tests, the findings with respect to response mode were consistent with those of Study 1: the superiority of the overt response over the covert and reading responses was definitely a function of the reproduction accuracy required on criterion items. With one negligible exception, this applied only to the items requiring reproduction of medical terms. While variations in frame size resulted in no significant difference among the groups for these same two tests, the variation of number of responses required per frame produced a significant effect in favor of multiple responses when test items required medical term responses. None of these effects were observed on the delayed retention test. Withholding confirmation produced no differential effect for any group. Moreover, there was no evidence of any interactive effects on achievement among treatments. Although program error rate was found to be a valid predictor of post-program performance, neither frame size, confirmation nor, with only one minor exception, number of required responses per frame resulted in any significant effects on the errors recorded for program performance. A significant effect found on program completion times was predictable in terms of the increase expected for written responses, expanded frame size and multiple response frame requirements. Response mode was also found to have a significant effect on the time it took subjects to complete each of the retention tests. Groups responding overtly on the program spent less time on the tests than groups responding covertly and, except on the delayed tests, than those using reading programs.

The overall results were discussed within the framework of verbal learning theory. The significant findings were interpreted by an analysis of the variables governing response learning, as opposed to associative learning, in self-instructional programs.

Contents

Chapter		Page
I	Introduction.	1
	Frame Content	1
	Response Mode	5
	Confirmation.	10
	Summary	12
II	The Relationship Between Response Mode and Response Difficulty in Programmed Instruction (Study 1). . . .	14
	Introduction.	14
	Instructional Materials	14
	Test Materials.	15
	Subjects.	15
	Procedure	15
	Results and Discussion.	16
III	Effect of Variation in Confirmation Procedures on Retention of Programmed Materials (Study 2)	19
	Introduction.	19
	Instructional Materials	21
	Test Materials.	23
	Subjects.	24
	Procedure	24
	Results	25
	Discussion.	26
IV	The Interactive Effects of Frame Content, Response and Confirmation Variations in Programmed Instruction (Study 3).	30
	Introduction.	30
	Subjects.	31
	Instructional Materials	32
	Test Materials.	40
	Procedure	45
	Results	46
	Retention Test Performance.	48
	Program Performance	61
	Test Completion Times	67
	Discussion.	68
References.		74

Figures

		Page
1	Distribution of Frame Lengths for the Basic Frame and Expanded Frame Programs.	34
2	Frame Sequence from the Multiple Response, Basic Frame Program.	36
3	Frame Sequence from the Single Response, Basic Frame Program.	38
4	Frame Sequence from the Basic Frame Reading Program.	39
5	Frame Sequence from the Multiple Response, Expanded Frame Program.	41
6	Frame Sequence from the Single Response, Expanded Frame Program.	42
7	Frame Sequence from the Expanded Frame Reading Program.	43
8	Differences in Mean Correct Response Percentages (Overt Response - Covert Response Test Scores and Overt Response - Reading Test Scores) for Two Retention Tests as a Function of Response Reproduction Accuracy.	53

Tables

		Page
I-1	Mean Number of Responses for the Various Response Accuracy Categories as a Function of Response Mode and Type of Retention Test Item	18
II-1	Program Performance: Mean Number of Correct Responses and Mean Completion Times (Minutes) for the Five Confirmation Procedures	25
II-2	Combined Daily Tests: Mean Number of Responses for the Various Response Accuracy Categories as a Function of Confirmation Procedure and Type of Retention Test Item.	27
II-3	Comprehensive Post-Program Test: Mean Number of Responses for the Various Response Accuracy Categories as a Function of Confirmation Procedure and Type of Retention Test Item.	27
II-4	Summary of Analyses of Variance of the Combined Daily Test Scores.	28
II-5	Summary of Analyses of Variance of the Comprehensive Post-Program Test Scores	28
II-6	Comprehensive Post-Program Test, Subtest V: Mean Number of Definition Test Responses for Two Response Accuracy Categories as a Function of Confirmation Procedure.	28
III-1	Percentage of Frames Requiring From 1 to 5 Responses in the Multiple Response, Basic Frame Program (384 frames)	33
III-2	Means and Standard Deviations of Program Completion Times in Minutes as a Function of Frame Size, Number of Responses Per Frame, Response Mode, and Confirmation	47

Tables Page

III-3 Means and Standard Deviations of Accurately Reproduced Definition Test Item Responses for Three Retention Tests as a Function of Frame Size, Number of Responses Per Frame, Response Mode and Confirmation Procedure. 51

III-4 Means and Standard Deviations of Correctly Recalled Definition Test Item Responses Scored Without Regard for Spelling Accuracy for Three Retention Tests as a Function of Frame Size, Number of Responses Per Frame, Response Mode, and Confirmation Procedure 51

III-5 Combined Daily Tests: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure 52

III-6 Combined Daily Tests: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Expanded Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure 52

III-7 Comprehensive Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure 54

III-8 Comprehensive Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Expanded Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure 54

III-9 Delayed Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure 55

III-10 Delayed Test: Mear Number of Medical Term Responses, by Response Accuracy Classification, for the Expanded Frame Programs as a Function of the Number of Responses Required Per Frame, Response Mode and Confirmation Procedure. 55

III-11 Summary of Analyses of Variance of Correctly Recalled and Accurately Reproduced Definition Test Item Responses for Three Retention Tests. 56

III-12 Summary of Analyses of Variance of Correctly Recalled Definition Test Item Responses Scored Without Regard for Spelling Accuracy for Three Retention Tests. 56

III-13 Summary of Analyses of Variance of Medical Term Test Item Responses by Response Accuracy Classification for the Combined Unit Tests 57

Tables		Page
III-14	Summary of Analyses of Variance of Medical Term Test Item Responses by Response Accuracy Classification for the Comprehensive Tests.	57
III-15	Summary of Analyses of Variance of Medical Term Test Item Responses by Response Accuracy Classification for the Delayed Tests.	59
III-16	Means and Standard Deviations of Correctly Recalled Definition Test Item Responses by the Basic and the Expanded Frame Program Subjects Scored with (Criterion I) and without Regard (Criterion II) for Spelling Accuracy for Three Retention Tests. . .	59
III-17	Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic and Expanded Frame Reading Programs on Three Retention Tests.	60
III-18	Means and Standard Deviations of the Types of Program Errors Made in the Frame Responses Elicited in Common by the Single and Multiple Response Frame Programs as a Function of Frame Size and Confirmation Procedure.	60
III-19	Summary of the Analyses of Variance of the Types of Program Errors in the Frame Responses Elicited in Common by the Single and Multiple Response Frame Programs.	63
III-20	Means and Standard Deviations of the Types of Program Errors Made by Subjects Instructed by the Multiple Response Frame Programs as a Function of Frame Size and Confirmation Procedure.	63
III-21	Summary of the Analyses of Variance of the Types of Program Errors Made by Subjects Instructed by the Multiple Response Frame Programs.	64
III-22	Product-Moment Correlations Between Program Errors (Incorrect and Incomplete Responses and Omissions) and Retention Test Performance	64
III-23	Product-Moment Correlations Between Program Errors (1 to 5-Letter Spelling Inaccuracy Medical Term Responses and Correct Definitions Containing Misspelled Words) and Retention Test Performance . .	65
III-24	Means and Standard Deviations of Test Completion Times in Minutes for Three Retention Tests as a Function of the Program Variations in Frame Size, Number of Responses Per Frame, Response Mode and Confirmation Procedure.	65
III-25	Summary of Analyses of Variance of Test Completion Times for Three Retention Tests.	66
III-26	Means and Standard Deviations of Test Completion Times in Minutes for the Subjects Instructed by the Basic and Expanded Frame Reading Programs on Three Retention Tests	66

I. Introduction

As pedogogical tools, linear self-instructional programs structure the material to be learned in small ordered units which are designed to guide the student gradually to a specifically defined level of subject matter mastery. All students are required to follow the same instructional sequence throughout the program, but each student is allowed to proceed at his own pace.

When we consider the instructional strategy and prearrangement of subject matter in linear programs we find that an eliciting stimulus, a response requirement and a confirmation procedure are designed into each of the steps, or frames, comprising the program. Accordingly, a frame containing an incomplete statement, question, or problem represents the stimulus material; the word, phrase, or solution which completes the statement, answers the question, or solves the problem represents the required response; and confirmation of the correct word, phrase, or solution is used to provide knowledge of results for the response.

One of the primary tasks in programmed instruction is to utilize techniques of frame construction that will optimize the association between the contextual frame material and the learner's constructed responses. Since the individual frame is the functional unit of any program, much of the program's effectiveness is dependent upon frame design. Research findings, however, have frequently indicated that the variables which currently characterize linear self-instructional frames are not any more effective in promoting associative learning than conventional methods of instruction. Moreover, there are elements of frame design which may lead to weak or inappropriate associations.

In the following, it shall be demonstrated that research on three of the most basic aspects of programmed instruction has raised a number of questions regarding the adequacy of the assumptions underlying the preparation of programs. The review is concerned with: (1) frame content variables such as frame size and the number of required responses within a frame, (2) the type of response required by the frame, and (3) the method of providing the student with confirming information. Only results based upon investigations using linear programs are considered.

Frame Content

Beginning with the early formulations by Skinner (1958, 1959), questions of frame content, such as the amount of information within a frame and number of responses per frame, have been the concern of investigators as well as of program developers. While there was general agreement on achieving subject matter mastery through a succession of small rather than large steps for guidance in developing linear programs, the meaning of "step size" depended upon each respective programmer's preconceptions.

The most common and simplistic approach to answering the step size question was to provide a series of short and easily followed segments designed to reach a specified behavioral objective. One of the initial tasks of the programmer was to systematically determine beforehand the precise set of examples and opportunities for practice that was considered necessary to teach the various concepts in the program. This usually involved a preconceived judgment concerning the number of frames required per concept and the number of responses that had to be made before the learner could reach the criterion behavior. Optimum frame size and content was then controlled through an empirical record of program performance. If a prescribed error rate, conventionally set between 5% and 10%, was exceeded in field-test tryouts, step size was altered to obviate the sources of difficulty. This was accomplished by the construction of more frames to clear up problem sequences, and/or through further explication of individual frame content. The program was then revised to insure that frame sequences could guide the learner gradually, and without errors, to a given level of proficiency.

Although error rate analysis and subsequent program revision seemed to be a feasible way of structuring the constituent elements of a program, questions regarding basic frame content still remained unresolved. Program errors were found to be too dependent upon such variables as the amount of prompting within frames to be considered as reliable criteria for determining step size. Consequently, there was no way, a priori, to guarantee that reducing program errors would increase program effectiveness.

One predominant feature of programs is their verbal content. Reading, then, is superimposed as an additional task the learner has to perform. The amount of reading material that any frame is allowed to contain establishes the basis for another attempt to define step size. This constraint has been variously interpreted as a prescribed number of ideas, facts, words or sentences per frame. As the state of the art progressed, as well as it being congruent with certain research findings (Kemp and Holland, 1966), programmers advocated limiting each frame to the content upon which the required response was specifically contingent. The rationale for defining frame content and consequently step size on this basis is derived from operant conditioning principles, and relates both to the contiguity of stimulus and response, and the frequency and temporality of reinforcement.

It may be seen that the interpretation of step size is limited either to the question of response difficulty, or to the magnitude of the behavioral increments - the number of successive approximations required for any single response to be elicited from a student, in the terms of a behavioral scientist. Briggs (1968) summarizes the various interpretations of step size as follows:

> (a) how difficult a response is to make, (b) how large a reading segment is presented before a response is required, (c) how much progress toward the goal is represented by one frame, (d) how long it takes the learner to make a response, (e) whether or not the student responds correctly, and (f) how frequently reinforcement occurs. (pp. 165-166).

Problems in applying such definitions to experimental situations are illustrated by a number of studies comparing the effectiveness of different step sizes (Gagne and Smith, 1962; Gagne and Bassler, 1963; Hamilton and Porteus, 1965; Shay, 1961; Smith and Moore, 1962). Gagne and Bassler (1963) and Hamilton and Porteus (1965), for example, report that the use of a greater number and variety of examples in their programmed material resulted in significantly better post-program retention. They interpreted these findings as support for the superiority of small steps. With each example constructed as a small step, those versions of the program which characterized a wide variety of responses were regarded as small step sequences.

In two other studies (Coulson and Silberman, 1960; Evans, Glaser and Homme, 1960) comparing the effects on retention of short and large step versions of the same program, the investigators created their separate versions by removing what they judged to be redundant frames. Coulson and Silberman extracted enough redundancies from a 104-frame unit on psychology to create a 56-frame unit. The former unit represented the small step version and the latter the large step version. Evans, Glaser and Homme followed essentially the same procedure. They began with a 51-frame program on mathematics and, reducing it progressively through the removal of "repetitive and transitional material," created 40- and 30-frame large step versions of the same program. Further, they constructed a 68-frame, or smaller step, version of the mathematics program by adding items to the basic 51-frame sequence.

Both studies demonstrated that the small step sequences, represented by those programs with the larger number of frames, produced significantly better post-test performance but only when the larger steps were produced through a reduction of the original frame material. Consequently, the superiority of the small step versions in these studies, as well as those previously cited, can be attributed to the inequality or paucity of material in the shorter programs rather than to the question of response difficulty or the magnitude of the behavioral increment. The results are notably different for the smaller step sequence represented by Evans, Glaser and Homme's 68-frame version of the program which was created by adding redundant material. This version, paradoxically, did not produce performance significantly different from the 51-frame version. The scores, in fact, appeared to indicate inferior results for this small step program.

Studies comparing the effects of small versus large steps suffer either from an indeterminant definition of step size or from the methodological issue of inequality among versions with respect to the subject-matter content. Simply elaborating on the number and variety of frames to decrease step size or, as in the last two experiments cited, extracting redundancies to increase step size, raises more questions than it provides answers.

A further consideration dealing with the matter of frame design concerns the number of responses learners are required to make in each program frame. Generally, they are found to vary from 1 to 5 in number. Aside from the obvious rule of thumb that too many blanks may result in ambiguous frames, programmers have not been provided with any empirical

research data to guide their efforts. While on the face of it, the problem of the number of responses required by a frame appears to be one of mechanics, under some conditions certain basic learning principles may be involved.

Before discussing these principles, let us illustrate such conditions by considering the stimulus context of the following frame:

> Glands that secrete hormones directly into the bloodstream are called _____ glands.

Conceivably, the same statement could have required a response to be made in the middle of the sentence or at two or more different positions in the same sentence. For example:

> Glands that secrete hormones directly into the _____ are called endocrine glands.
>
> Glands that secrete hormones directly into the _____ are called _____ glands.
>
> Glands that secrete _____ directly into the _____ are called _____ glands.

At given stages during the program the learner could be expected to provide correct responses to any of these statements. In fact, any one of the four versions cited exemplifies routine occurrences in the composition of frame items.

It is not difficult to see that in order to complete the latter three frames the student would probably be forced to read past the blanks, thereby retracing his steps in a discrete fashion one, two or three times, to make the appropriate responses. In order to respond to the first blank in the last frame, for example, the student requires additional information and has to read to the end of the frame, then back to the middle, and finally back to the beginning. In an attempt to fill in the response blanks, the learner is unable to read the material in the syntactic order presented in the frame, being forced rather into an erratic reading and responding sequence. This may result in either a loss in association of the learner's response with the stimulus material to which he is responding, or in associations other than those intended by the programmer. The process of associating spatially remote items in a series as a consequence of seeing them contiguously or sequentially has been experimentally demonstrated in terms of the formation of remote forward and backward erroneous associations (McGeoch and Irion, 1952). Of further importance to this analysis are early findings by Hall (1928) and Lepley (1934). These investigators found that although the presence of remote associations may not be evidenced immediately after learning, a considerable amount may appear in tests given at some time after the original learning. Thus, while the learner is able to respond to program frames correctly, we may infer that his ability to accurately retain the associations learned for any long period of time may be impaired.

When these considerations are applied to the previously illustrated

frames, the learner ultimately will be unable to recall the relationship existing between the secretion of hormones, the bloodstream and the endocrine glands, especially if the intended associations are weakened because of a loss of contact between a response and its proper context. On the other hand, if erroneous remote associations occur during learning, the stimulus context may elicit a response which the programmer never intended as a direct association, such as having the learner recall that "hormone glands secrete endocrine."

Response Mode

One of the basic principles underlying self-instructional programming is the requirement that the learner respond overtly to the subject matter. This is implemented, primarily, by the constructed response.* The literature is replete with studies comparing equivalent versions of a program in which one group is required to respond overtly and a second group which is instructed to respond covertly. In the covert response mode, the subject "thinks" of the answer that completes the statement.

Cummings and Goldstein (1964) reported results which clearly indicated a significant difference between written and covert program responses. Using a 119-frame program on the diagnosis of myocardial infarction with student nurses and technicians, these investigators found that overt responding was statistically superior to covert responding on both an immediate and a delayed retention test. A variety of other studies, however, have failed to support such a positive finding. The following is a list of studies reporting no significant differences between overt and covert responding:

a. Evans, Glaser and Homme (1960a) investigated differences between the two response modes with undergraduate psychology students. The subjects studied a program on the fundamentals of music. Other specific data concerning the program or criterion test were not reported.

b. Evans, Glaser and Homme (1960) again found that overt and covert responding on a 72-frame symbolic logic program had no effect on achievement. The investigators used three types of criterion tests: true-false, short answer recall and problem solution.

c. Hughes (1962) administered a 719-frame programmed text on IBM data processing to industrial trainees. The criterion test consisted of multiple-choice items.

---

* For the purposes of this review, the overt or constructed response is defined as the word part, word or phrase which is represented by a blank in a frame and must be written out in its entirety in order to complete a statement or answer a question. Programs which require responses such a writing a matching letter or number, manipulating a multiple-choice button, underlining correct alternatives or answering audibly, will be generally omitted from consideration.

d. Lambert, Miller and Wiley (1962) used an 843-frame programmed text on methematical sets, relations and functions. Ninth grade students were compared at three levels of mental ability. The characteristics of the criterion test were not reported.

e. Stolurow and Walker (1962) administered a 60-frame program on descriptive statistics to a combination of psychology students, education majors and elementary school teachers. A variety of tests were used to measure immediate and delayed (two weeks) retention.

f. Crist (1966) used two commercially available programs, one on latitude and longitude (351 frames) and the other on the solar system (331 frames), with sixth grade students. The items appearing on an immediate and a six week retention test were not described.

g. Yarmey (1964) used a 343-frame primer on programmed instruction. Undergraduate psychology students served as subjects. A short answer recall test was administered immediately and four weeks after completion of the program.

h. Wittrock (1963) used a 280-frame tape and slide program on kinetic molecular theory with first and second grade students. The immediate and delayed test (one year) consisted of multiple-choice items.

The issue of overt responding, as it is implemented generally in classroom learning and particularly as it is designed into programmed instruction, has been the subject of a number of reviews (Holland, 1965; Lumsdaine and May, 1965; May, 1966; and Anderson, 1967). In addition to comparisons between overt and covert responding, these reviewers considered research findings from studies utilizing instructional materials in which there was no provision for responding. In programmed instruction research, materials which required nothing more than reading each frame in its completed form were originally used to provide criterion test data against which the effects of overt and covert response modes could be compared. The introduction of these "reading" programs, however, raised further questions regarding the value of the constructed response.

In one study using a 35-frame program composed of discrete factual items concerning men of historical note, geographical information, etc. Goldbeck and Campbell (1962) compared overt, covert and reading programs with seventh grade students as subjects. These investigators found that retention test scores were a function of response mode interacting with the amount of cueing. With maximal cueing (20% error rate), the reading and covert programs resulted in retention test scores significantly higher than the overt program. At a moderate level of cueing (57% error rate), the scores of the overt program were significantly higher than the covert group with the reading program scores falling at the intermediate level. When the program frames were miniminally cued (80% error rate), the highest retention test scores were experienced with the reading program. Scores for the covert program were next highest and the scores for the overt program were lowest. For minimal cueing, however, the differences were not statistically significant.

Alter and Silverman (1962) compared reading and overt response programs under different conditions of program presentation: machine vs. text administration and self vs. external pacing. An 87-frame program on basic electricity was used for the machine vs. text comparison, while a 90-frame program on binary numbers was used to determine the effects of pacing. College undergraduates served as subjects. In almost all of the comparisons the reading programs represented the source of fewest errors on retention tests, which were composed of written response and multiple-choice items. Statistical analyses did not yield any significant differences except in one condition, and in this instance reading was superior.

Feldman (1965) studied the effects of covert responding and reading with two groups of college sophomores differentiated into low and high verbal ability on the basis of SCAT scores. An Introductory Psychology program was used for the study. No significant differences were found between covert and reading subjects on pre/post-test gain score comparisons. When analyzing the criterion test scores alone, however, Feldman found a significant difference in favor of the reading program for subjects in the low verbal ability group.

Findings inconsistent with those cited above have also been reported. An experiment by Krumboltz and Weisman (1962), for example, provided results comparing the effects of overt responding, covert responding and reading on the retention of material in a 177-frame program on statistical analysis. While an immediate posttest did not reveal any significant differences among the response modes, an alternate test given two weeks later proved otherwise. The overt response mode resulted in significantly higher retention scores, with the reading and covert response mode scores being almost identical.

Jacobs, Yeager and Tilford (1966) compared overt responding and reading using a 300-frame program on the Bill of Rights. Eleventh grade students served as subjects. Multiple-choice tests measuring the retention of both factual and conceptual knowledge were administered immediately following the program and again six weeks later. The overt responding subjects scored higher than the reading subjects on both the immediate and delayed criterion measures. Only the differences on the immediate test, however, were statistically reliable.

Barlow (1967) also found a significant difference in favor of an overt response group when retention was measured on a multiple-choice test. Using freshmen psychology students and a 480-frame program concerned with the essentials of learning theory, Barlow not only compared overt and reading programs but also related the test results of the subjects in the two response mode groups to SAT scores. The difference in test scores between overt and reading subjects with SAT scores lower than 500 was almost twice as great as the difference observed when the response mode comparison involved subjects with SAT scores higher than 500.

Grace and Cantor (1966) presented a program on the drug control of alcoholism to alcoholic VA hospital patients. The number of variables being investigated (overt vs. reading, sequenced vs. scrambled material,

immediate confirmation vs. optional confirmation vs. no confirmation) tends to obscure the role of response mode in this study, but post-program performance for overt response subjects was superior to that of reading subjects for two out of the three significant differences reported.

A substantial number of studies have been unable to find any significant differences between reading and either overt or covert responding. The following is a list of such investigations:

a. Feldhusen and Birt (1962) compared overt responding and reading on a 39-frame program concerned with the principles of programmed instruction. The programs were administered to students enrolled in an undergraduate course in general psychology and a retention test was given immediately after the program. The types of items included in the retention test were not specified.

b. Fiks (1964) used three different programs to compare the effects of reading, overt and covert responding. The subject matter areas investigated were the concept of reinforcement (20 frames), weightlessness and space travel (24 frames), and the value of automobile seat belts (24 frames). Visitors at a state fair were used as subjects, and tested with a true-false immediate retention test.

c. Hartman, Morrison and Carlson (1963) compared reading and overt responding on a 1756-frame program on IBM machine operation. The program which consisted of units on IBM cards, card punchers, sorters and reproducers was given to customer trainees. The composition of the post-program test was not specified.

d. Reid and Taylor (1965) compared reading with overt responding with a 580-frame program on the process of papermaking using undergraduate students as subjects. Written response tests were given immediately, and 12 weeks after, the completion of the program.

e. Roe (1962) compared reading and overt responding. A 192-frame program on elementary probability theory was administered to students enrolled in a freshman engineering course. No information was provided about the items in an immediate retention test.

f. Tobias and Weiner (1963) provided a comparison of reading, overt and covert response modes on a 90-frame program concerned with the addition of binary numbers. The program was administered to undergraduate education majors. Short answer completion items were given immediately, and 6 weeks after program completion.

g. Warren (1966) compared reading and overt responding. An 85-frame program on the British currency system was given to the overt response group. The reading group was provided with a table of information containing British currency equivalents.

A multiple-choice and a written response test was administered after the instructional session.

A number of studies comparing response modes all used sets from the same program, The Analysis of Behavior (Holland and Skinner, 1961). Holland (1960), using sets 9-21, found posttest performance favoring an overt response in comparison with reading. Williams (1963), with sets 7-11, reported similar findings. This would seem to indicate that the Holland-Skinner program contains features which benefit from the written response. Findings from two other investigations, however, do not support this supposition. Gilpin, cited by Barlow (1961), used sets 1-8 of the same program and found no significant differences between similar response mode groups. Another study using sets 1 and 2, carried out by Stewart and Chown (1965), investigated reading and overt response modes with female volunteer subjects divided into two age groups: old (51-59 years) and young (20-36 years). Old and young subjects using the reading program performed significantly better on the criterion test than old subjects responding overtly. However, no difference in performance was observed between young-reading and young-overt groups.

Williams (1963) presents evidence indicating that the inconsistencies among experimental findings may be due to differences in the composition of criterion test items. In her study a constructed response version of the program was compared not only with a standard reading program, but with one in which the filled-in responses were underlined for emphasis. A fourth version which required multiple-choice responses was also used. The results of a 20-item objective test administered to undergraduate students indicated that the criterion performance of the constructed response group was significantly superior to either of the reading groups. Further analysis disclosed that there was no significant difference between constructed response and multiple-choice groups when compared on the basis of overall group means. However, the difference between the groups was decidedly significant when the comparison was based exclusively on those test items measuring the subject's performance on the novel technical terms he had studied in the program.

An additional study by Williams (1965) using a different program with younger subjects provides further information on the relationship between response mode and test item difficulty. A 120-frame program covering the scientific classification of animals was administered to sixth grade pupils, and a comparison was made between constructed and multiple-choice responding. A retention test composed of 16 written response items and 16 multiple-choice items was given on the day following program completion. Each subtest was equally divided into items which required the use of technical terms taught by the program, and items which could have been answered with terms already in the vocabulary of sixth grade students. The constructed response mode was found to be superior to the multiple-choice mode only on criterion test items which required the written reproduction of the complex technical terms studied in the program.

In summary, the alleged merits of the written response appear not to have been experimentally resolved. Reading programs have been frequently found to be as equally effective as constructed response

programs. Moreover, many of the investigators who have compared the various response modes emphasize that learning efficiency (retention test scores as a function of time taken to complete the program) is yet another factor in making judgments. A number have concluded that, regardless of the gains registered by the written response mode in some investigations, economic considerations still dictate the exclusive use of covert response or reading programs.

Confirmation

In standard linear programs, the learner is instructed to answer the question posed by the material within a frame, or to fill in the deleted part of a statement, and then to compare his response with the correct one. Knowledge of the correct response is made available as an integral part of the program, and is generally concealed until the learner's response is made. Allowing the learner to determine immediately the adequacy of his response has been called confirmation. Confirmation of the correct response was posited as the functional counterpart of reinforcement in animal conditioning and, as such, was considered indispensible in programmed learning. However, the benefits to either program or post-program performance resulting from utilization of this feedback procedure have not been experimentally verified.

Meyer (1960) provided some early promising results concerning the role of confirmation. Eighth grade students were presented with programmed booklets designed to teach the derivation of English words through a knowledge of commonly used prefixes. She reported that students receiving immediate knowledge of the correct response obtained significantly higher pretest-posttest gain scores than students working with booklets which did not contain any confirming information.

In an attempt to prevent the copying of answers, which is possible when programmed materials are presented in booklet form, Moore and Smith (1961) administered a spelling program to sixth grade students on a teaching machine. These investigators found that students who were prevented from seeing the correct answer while going through the program on the teaching device obtained higher scores on four unit tests and a comprehensive test than students who did not have their view of the correct responses obstructed. None of the differences, however, were statistically significant.

Additional experiments designed to compare the effects on posttest achievement scores of treatments represented by confirmation and non-confirmation groups have also failed to obtain significant differences. These include Holland (1960), who used a program on operant conditioning techniques; Feldhusen and Birt (1962), who used a short, 37-frame program on the principles underlying teaching machines and programmed instruction; and Hough and Revsin (1963), who used a selected response, 355-frame linear program on the history of secondary school education.

Ripple (1963) found that undergraduates who were taught the principles of programmed instruction by a 134-frame programmed textbook devoid of confirmation material made almost twice as many errors while

learning as a group using the text complete with correct response information. The presence or absence of confirmation during the program, however, was not reflected in posttest performance scores. An analysis of retention tests administered two days and ten days after completion of the instructional materials did not uncover any significant effects. Recognition as well as recall items were included in these tests and were analyzed independently. The absence of any differences between confirmation and non-confirmation groups was further supported by an additional finding demonstrating that both proved to be equally superior to a reading version of the same program.

At the onset program writers and experimental investigators alike regarded the use of confirmation as a reinforcement procedure. It was inevitable, then, that studies would eventually address themselves to the dimension of reinforcement schedules. Extrapolating from the effects of partial reinforcement on extinction in animal studies, it was hypothesized that omitting knowledge of the correct response intermittently would prolong the retention of programmed subject matter. Krumboltz and Weisman (1962), for example, used a 177-frame programmed textbook on educational measurement to obtain some information bearing upon this matter. They compared totally confirmed and non-confirmed versions of the program to four partially confirmed versions. In the four programs containing the partial confirmation, versions representing one-third or two-thirds of the frames being followed by the correct response were further subdivided into treatments that characterized confirmation for regular and irregular sequences of such frames. This provided confirmation schedules ranging from 0% to 100%, with two fixed-ratio schedules of 33% and 67%, and two variable-ratio schedules with comparable percentages of confirmation. These investigators found that the variations in amount of confirmation produced definite differential effects on program performance, but not on posttest achievement. It was determined that as the percentage of program frames followed by correct answers increased, the number of errors made in response to program frames decreased. The fixed and the variable schedules had similar effects.

Rosenstock, Moore, and Smith (1965) pointed out that constructing partially confirmed programs by randomly deleting correct answer material could confound a comparative evaluation of different schedules. They argued that a fixed and a variable ratio program containing the same number of confirmed response frames could still vary greatly on dimensions other than confirmation sequence. One of the programs, for example, could inadvertently provide a preponderance of correct response information for highly-cued practice frames. Subsequent retention test scores, then, would reflect not only differences in the confirmation schedules utilized, but also differences in the redundancy of the information supplied by the programs through confirmation. Discrepancies in the subject matter content of the confirmed frames would constitute an additional confounding effect. With such methodological considerations in mind, these investigators used a program on mathematical set theory. By carefully selecting comparable confirmation frames, they developed 20% fixed-ratio and 20% variable-ratio versions of the program. These intermittently confirmed programs, containing 330 frames, were administered to groups of sixth grade students along with programs con-

taining 0% and 100% confirmation. The results indicated that the manipulation of the confirmation schedules did not produce any differential achievement effects on tests given immediately after, and two weeks following, the completion of the programmed materials. An analysis of program responses revealed results consistent with those reported by Krumboltz and Weisman (1962). The continuous confirmation condition yielded significantly fewer program errors than the other treatment conditions. The authors, however, emphasized that the error findings should be cautiously interpreted since the use of a text for program administration did not preclude the possibility that copying may have influenced the error rate data.

A study by Lublin (1965) presented evidence that greater criterion test achievement was associated with less dependency on confirmation. Using three versions based on 27 sets of The Analysis of Behavior (Holland and Skinner, 1961), Lublin demonstrated that both no confirmation and a 50% variable-ratio schedule produced significantly better performance on an immediate posttest that did continuous confirmation. While not superior to that of the 50% variable-ratio program, the non-confirmation condition resulted in achievement significantly greater than that of the 50% fixed-ratio program. Interestingly enough, the no confirmation program required the most time for average completion and the continuous confirmation the least time.

Jacobs and Kulkarni (1966) offered additional evidence favoring the interpretation that the presence of confirmation may be associated with a decrement in achievement. A set of booklets was used to administer a 273-frame program in chemistry to high school students. For one group of students correct answers were provided on the reverse side of the booklet pages containing the frame material. The correct answers were omitted for another group of students. Significantly higher scores on a criterion test composed largely of multiple-choice items were achieved by the no confirmation group.

Summary

In each of the areas discussed, a review of significant studies has identified the specific issues which form the basis for the present series of investigations. The comparisons of studies dealing with frame content revealed inconsistencies in precisely defining this program variable. Although such studies have produced several significant findings, the varied and sometimes questionable approaches to manipulating frame content preclude generalizing definitive statements about this aspect of frame construction.

While the literature pertaining to response modes in a general learning context abounds with studies substantiating the value of overt responding, its contribution in programmed learning is still in doubt. Studies purporting to identify the conditions under which this mode is superior in programs are countered by the results of numerous others that find covert and reading response modes equally effective. The question remaining is whether response mode has a differential effect on the specific type of response being elicited.

The value of response confirmation, although long held to be indispensable to programmed learning, has been dispelled by virtually all of the studies cited. Only the question of other logical modifications in confirmation procedure and their possible effect on post-program achievement would still appear to be unanswered.

Three separate but closely related studies were designed to investigate the issues raised with respect to frame content, response mode and confirmation procedures. In the first study, the overt is compared with the covert response mode; in the second study, a variety of confirmation procedures, including non-confirmation, are compared for effects on post-program achievement; and in the third study, the interactive effects of response mode, confirmation procedures and carefully defined variations of frame content are measured in a factorial design. This last study also provides for a comparison of reading programs with other response mode programs. The detailed account of each study to follow will illustrate how the third and final study deliberately capitalized on the experiences and results of the earlier two in organizing the dimensions of the experimental design.

II. The Relationship Between Response Mode and Response Difficulty in Programmed Instruction (Study 1).

Introduction

Relatively few studies since 1966 have reportedly addressed themselves to the issue of overt versus covert responding in programmed instruction. Up to that time, the literature abounded with studies comparing the effects of these response modes. The experimental findings were largely inconsistent. Studies such as those performed by Evans, Glaser and Homme (1960) and Crist (1966) exemplify the many which found no difference in effectiveness between overt and covert responding; those by Holland (1960) and Cummings and Goldstein (1964) exemplify the fewer in number which found overt responding to be superior; and that by Silberman, Malargno and Coulson (1961) is among the very few which reported the superiority of the covert mode.

Williams (1963, 1965, 1966) has provided evidence indicating that the inconsistencies among experimental findings may be attributable to differences among programs in response difficulty, with the more difficult response favoring the overt mode. In all three of her studies no differences between response modes were observed when subjects could answer test items in general and familiar terms. Responding in writing, however, was found to be the superior response mode when retention test items required the recall of more difficult material such as, technical terms.

The present study represented an attempt to specify further the interactive effect of response mode and response difficulty on the retention of programmed subject matter. The research design involved the utilization of a linear program which allowed for the manipulation of two specific levels of response difficulty within the same program, and the subsequent determination of the effects of overt versus covert responding on retention test scores as a function of response difficulty.

Instructional Materials

The experimental program was a 197-frame, mimeographed modification of a portion of a commercially available linear program on medical terminology (Smith and Davis, 1963)*. The commercial program was developed to teach high school graduates the recognition and reproduction of medical words through a knowledge of prefixes, suffixes, word roots and combining forms derived from Greek and Latin words.

---

* The investigators are grateful to the publisher, John Wiley and Sons, Inc., and the authors, Genevieve Love Smith and Phyllis E. Davis, for permission to use and modify portions of the program entitled Medical Terminology: A Programmed Text, 1963.

The first 28 frames in the experimental program provided a review of basic word parts and their use in building compound words. The remaining frames were concerned with teaching 44 medical word parts which were used to build 55 different medical words. The contextual material in the vast majority of the last 169 frames elicited, as required responses, either the definition of medical words and word parts or the reproduction of these medical words and word parts. Each frame called for from one to five responses.

The selection of medical terminology as the programmed subject matter was based upon two considerations. First, it was assumed that the technical nature of the vocabulary would control to a large extent the degree of prior familiarity with the responses required by the program, and yet deal with a subject matter area interesting enough to enlist and maintain the cooperation of the experimental subjects. Secondly, it was assumed that responses requiring the reproduction of medical words and responses requiring the definition or meaning of medical words would constitute two distinct levels of response difficulty. The reproduction of medical words requires that the learner initially respond with unfamiliar technical terms, and this, it was postulated, would be a more difficult response to make than one in which the recognition of medical terms calls for definitions containing words already present in the learner's vocabulary. Thus, when medical terms such as gastrectomy and adenoma are presented to subjects in frames or test items, the contention was that responding with definitions for these terms would not be as difficult as supplying the actual medical terms corresponding to stomach excision and glandular tumor.

Test Materials

The retention test contained 120 items. Half of the items measured proficiency in the recall of either medical words or medical word parts when given their definitions; the remaining items were concerned with the recall of definitions, given their corresponding medical words or word parts. The responses required by the test items were identical to those elicited by the program.

Subjects

A total of 50 unpaid volunteer undergraduate students from Northeastern University, Boston, voluntarily participated in the study. Students enrolled in premedical and medically allied programs were excluded. Each volunteer was questioned to ascertain that his knowledge of medical terminology was minimal. Subjects were randomly assigned either to an overt or to a covert response group. Two subjects from the covert response group were unable to attend the testing session. Consequently this provided a sample size of 26 for the overt group and 24 for the covert group.

Procedure

A TMI-Grolier Min/Max III Teaching Machine was used to present

the programs to individual subjects. Each reported for an independent session and no more than four subjects were allowed to participate at the same time. The subjects were given approximately half of the 197 frames on the first day and completed the remaining frames on the following day. The overt response group was instructed to complete the frame material by making written responses, while the covert response group was told to think about the responses that would best fill in the blanks appearing in the frame material. The average time to complete each session was approximately one hour.

The retention test was administered seven days after the completion of the program. The possibility of subjects being aided in their recall of specific test items which would have resulted from an exposure to other items in the test, was minimized by administering the test in the same teaching machine that was used for program presentation. Subjects were prevented through mechanical safeguards from seeing more than one item at a time and from going back to previously presented items.

Each subject was instructed to spell all of his answers to the best of his ability, and to use, as accurately as possible, the wording of the program as a source for his answers to the test items.

Results and Discussion

To evaluate the results of the experiment, the investigators had to define a continuum along which various categories could be established for analyzing the accuracy of the criterion test responses. A large number of the responses elicited by the test items could not be scored by just a simple correct or incorrect designation. Misspelled medical terms posed a particular problem. Some of the medical terms were misspelled but were still identifiable as the required test item responses. Spelling inaccuracies in other instances rendered the terms as either unrecognizable or indicated possible confusions with similarly spelled medical terms in the program. Consequently, different response accuracy categories were used in evaluating the results of the test items eliciting medical terms. These were:

1. Term is accurately spelled.

2. Term is incorrectly spelled: one letter is either incorrect, added, omitted or transposed. The misspelling does not change the meaning of the medical word or indicate a confusion with another medical term in the program.

3. Term is incorrectly spelled: two letters are incorrect, added, omitted or transposed with the restrictions observed in category 2.

Determining whether the required response was still recognizable when more than two letters were involved in the misspelling proved to be unreliable. This was especially true in analyzing the short medical word parts.

Unlike the procedure used for medical terms, spelling inaccuracies were not considered in categorizing the test items requiring definitions of medical terms. Rather, deviations from the wording utilized in the program provided the basis for scoring these responses. The following categories were used:

1. Subject responded with definition used in the program.

2. Subject did not use the exact definition used in the program. The definition, however, conveys the same essential meaning and does not indicate a confusion with another medical term or definition in the program.

Any responses, whether medical terms or definitions, falling outside of these categories, and any omissions, were scored as incorrect. Three judges were used to categorize all of the criterion test responses.

The mean retention test scores are presented in Table I-1 as a function of response mode and type of retention test item. The scores presented in this table are entered according to the response accuracy categories described above. The entries are cumulative from top to bottom. The "0 to 2-letter spelling inaccuracy" category for the test items requiring medical terms, for example, includes the entries of the "accurately reproduced" and "1-letter spelling inaccuracy" categories in addition to its own contribution to the total entry. The mean score listed in the "acceptable approximation" category for the definition test items contains accurately worded definitions as well as those deviating from the wording used in the program.

As can be seen from the table, the overt response group obtained higher retention test scores than the covert group on both types of retention test items. It is of special importance to note that for the test items requiring the recall of medical terms, the difference between overt and covert response groups is the largest in the "accurately reproduced" category and becomes progressively smaller as response accuracy decreases. In addition, note that the difference between response mode groups is the smallest for the test items requiring the definitions of medical terms.

A rectangular distribution of the covert response group's retention test scores along with extreme variability in the scores of both groups indicated a need for a nonparametric statistical evaluation of the results. A Mann-Whitney U test revealed that the differences observed between the two response mode groups were not significant when considering the test items requiring the definitions of medical terms. With the medical term recall test items, however, the statistical evaluation indicated superior retention by the overt response group only when responses without spelling errors were compared. The difference of 5.1 observed between the two groups in the "accurately reproduced" category was found to be significant at the .05 level with a one-tailed test. It can be observed in Table I-1 that as medical term response accuracy decreases the differences

TABLE I-1

Mean Number of Responses for the Various Response Accuracy Categories as a Function of Response Mode and Type of Retention Test Item

Retention Test Items	Response Accuracy Categories	Response Mode		Mann-Whitney U Analysis (z)
		Overt	Covert	
Medical Terms	Accurately Reproduced	27.5	22.4	1.67*
	0 to 1-Letter Spelling Inaccuracy	30.8	26.3	1.25
	0 to 2-Letter Spelling Inaccuracy	32.0	28.0	1.05
Definitions of Medical Terms	Accurately Worded	36.4	33.4	0.95
	Accurately Worded and Acceptable Approximation	37.5	34.1	0.97

* $p < .05$

between the response mode groups, as well as the size of the values associated with the Mann-Whitney U analysis, also show corresponding decreases.

In conclusion, under the conditions that prevailed in the study these findings indicate that an interactive effect between response mode and response difficulty governs proficiency in the recall of programmed material. The retention of subject matter is more effective with a program requiring an overtly constructed response only when two conditions are fulfilled: 1) the retention test must elicit the recall of relatively difficult material, and 2) the retention test scoring procedure must require the accurate reproduction of that material.

III. Effect of Variation in Confirmation Procedures on Retention of Programmed Materials (Study 2).

Introduction

In spite of isolated instances, such as Meyer (1960), where confirmation appeared to have a positive effect on posttest achievement, the mass of contrary evidence raises serious doubts concerning the role of confirmation in programmed instruction (Holland, 1960; Moore and Smith, 1961; Feldhusen and Birt, 1962; Krumboltz and Weisman, 1962; Hough and Revsin, 1963; Ripple, 1963; and Rosenstock, Moore and Smith, 1965). Moreover, additional evidence indicates that the practice of confirming responses in programs may be responsible for a decremental effect on achievement (Lublin, 1965; Jacobs and Kulkarni, 1966). The authors of the present study proposed that if confirmation per se does not function properly in its alleged role as a reinforcing agent in programmed instruction, its ineffectiveness may be attributed to the way in which responses are conventionally confirmed.

Citing an example of hypothetical frames designed to teach the application of the rule "i before e except after c" will help to illustrate a potential inadequacy of techniques ordinarily employed by programmers to provide knowledge of results. Suppose, instead of having a student spell an entire word, the programmer allowed the missing letters (like those in BEL__VE and REC__VE) to be practiced in isolation. Further, suppose that after each required response, the programmer confirmed only the actual response, that is, the letters practiced, rather than the response as part of the whole word. In his spelling repertoire, then, the student may very well master the responses IE and EI, but still experience difficulty with a request to spell DECEIVE.

Some self-instructional programs require the learner to construct his responses within the context of the entire frame. Where this is the case, the spelling analogy does not fully apply. It does, however, have direct relevance in programs where the learner is required to write his responses on separate answer pads, etc. Of even more importance, and generally applicable to all self-instructional programs is the fact that the programming procedure allows the response term to be confirmed in isolation. That is, after the learner fills in the blank in a frame, the confirmation item reveals only the correct answer apart and severed from the frame context. It is comparable to asking the student to fill in the missing letters in PERC__VE, REL__VE, and CONC__VE and then confirming his responses by showing him EI, IE, and EI.

When a response is confirmed in isolation, i.e., without a restatement of the appropriate eliciting frame material, an inevitable time lapse occurs between the instant the learner decides upon his response and the exposure of the correct response. During this interval the learner may experience difficulty in spelling the response word, he

may think of an alternative response, or he may engage in any number of activities through thought or action that are alien to the programmed subject matter. In any event, this time lapse possibly permits the intervention of extraneous or distracting events which can interfere with what the learner responded to and the actual response. Accordingly, a student may know that his response was correct, or incorrect, and not remember what elicited the response.

By way of tacitly acknowledging this possible decremental situation, some programmers include instructions with their programs suggesting that the student re-read the completed frame after responses have been made. Although this would seem to be a satisfactory method of recovering any loss that may have been incurred through a disruption of the chain of associations in the frame, it is doubtful that many students would consistently follow these instructions in preference to proceeding immediately to the next frame. Additionally, re-reading the frame after an incorrect response is made and before reviewing the confirmation item would result in the student's exposure to misleading information.

Presupposing the validity of the above analysis, it appears plausible that if confirmation has any particular merits its positive contributions are possibly nullified by a procedure that dissociates the learner's response from its eliciting material. One way of eliminating the undesirable aspects of this feedback process would be to have the confirmation information include the eliciting frame material as well as the correct response. This would rejoin, so to speak, the associations intended by the programmer, and it would make the method of confirmation more consistent with learning principles which advocate practicing a response in the presence of its appropriate stimuli (Guthrie, 1952; Estes, 1960). The effectiveness of this type of confirmational procedure, which can be designated as confirmation in context, has been examined by Krumboltz and Bonawitz (1962). They used a 153-frame program on principles of achievement test construction to test the hypothesis that programmed learning would be more effective when responses are confirmed in context than when they are confirmed in isolation. It was found that although knowledge of terminology learned yielded insignificant differences, the context group did score significantly higher on retention test items which measured the application of the principles learned. Two aspects of their experimental conditions, however, indicate that the results may be inconclusive. First, the specific response words appearing in the confirmational material were underlined and hence conspicuous even within context. When questioned, subjects admitted that they frequently did pick out the underlined response as though it were in isolation. Second, the experiment did not include a comparison with a non-confirmation condition.

An unpublished study by the first two authors of the present report provided a comparison of the effects of confirmation in isolation and in context with those of a reading program with no confirmation. These investigators used a linear program consisting of 83 frames and 10 panels on the topic of light and characteristics of lenses. Five different versions of the program were administered to 135 college sophomores in an introductory psychology course. These five versions were: (1) overt response confirmed in isolation, (2) covert response confirmed in isolation, (3) overt response confirmed in context,

(4) covert response confirmed in context, and (5) a reading program. A multiple-choice test administered immediately after the instructional session did not reveal any significant differences among the programs. A second test with parallel items was then given six weeks later. In the second test 48% of the students in the overt group and 50% of the students in the covert group who received confirmation in context achieved a score of 65% correct or higher. The percentages of students attaining comparable scores in the overt and covert groups confirmed in isolation, and in the reading group, were 33%, 31% and 32%, respectively. Other cut-off scores in the vicinity of 65% resulted in essentially the same differentiation among groups.

These preliminary findings suggested that the method of confirming responses in context warranted further research and provided the purpose for the present investigation. One consideration in the present investigation involved the effectiveness of feeding the correct response back to the learner, either inconspicuously or highlighted, along with that portion of the frame which elicited the response. Two additional comparisons were concerned with conventional confirmation methods. With the inclusion of a no confirmation treatment, the investigation provided an evaluation of a total of five variations in response confirmation.

Instructional Materials

The investigators prepared a 378-frame linear program on medical terminology adapted from portions of the programmed text by Smith and Davis (1963). The first part of the program had been used earlier as the experimental program for Study 1. It was revised prior to its use in the present study to eliminate common sources of difficulty. All frames with an error rate greater than 5% were rewritten or deleted. Additional frames had to be written in some instances to clear up problem sequences. The number of revisions turned out to be minimal.

As in the first study, the program was designed to teach students how to recognize and build medical words from a knowledge of their component word parts. The first 30 frames were used to provide a review of word building principles. Throughout the remainder of the frames, 96 medical word parts were introduced and concurrently used to form 155 medical terms. The program frames were designed to train students to recall and define medical word parts and the medical words they formed.

Five mimeographed versions of the program were prepared to accommodate the various treatment conditions. Essentially, the same program was used for each treatment except for a specific variation in response confirmation. A TMI-Grolier Min/Max III Teaching Machine was used to administer all versions of the program. The machine prevents subjects from reviewing any previously exposed frames and consequently from changing any responses after a program segment has been advanced.

One of the experimental treatments may be described as a procedure conventionally used in programmed instruction, especially when a teaching machine is involved. The subject reads a small segment of the material framed by a plastic window on the teaching device. He writes his response, consisting of single terms or short phrases, through a

cut-out in the plastic along the lower part of the frame. By manipulating a wheel, he moves past the framed material leaving only the responses he made to be compared with the newly exposed correct response information. All other parts of the original frame have disappeared from view. The subject's responses cannot be altered since they have been positioned behind the plastic window and the machine cannot be reversed. With this procedure, a response is confirmed in isolation of its original frame context and displayed in absence of the frame material.

The second confirmation procedure is also quite common in programmed instruction, particularly where the material is presented in booklet form. When using this type of program, the student is instructed to keep the confirmation masked until the required response is made. After the mask is removed, the student's response, the correct response and the eliciting frame can be viewed simultaneously. This procedure, consequently, allows the response to be confirmed in isolation, but the frame material is available for review.*

This second confirmation variation was made possible on the teaching machine by enlarging the frame window of the device. The modification kept the frame material as well as the subject's response in view after the program was advanced to expose the correct response.

The third variation provides for the possibility that conventional modes of confirmation, such as represented by the first procedure mentioned, succeed in weakening the association between the response and its eliciting frame material. In this third treatment, the subject writes his response through the cut-out in the plastic window and moves the study frame up and out of view to expose the confirmation. However, confirmation in this case includes not only the single response but the original frame context as well. If the original is a short frame, it is fed back in its entirety; if a long frame, only the vital part of the context is included in the confirmation. Whenever a frame requires multiple responses, all parts of the frame are fed back regardless of relevancy. The actual response terms within the context are not made conspicuous in any way. The treatment represented by this variation, in which the response is confirmed in context, requires subjects to attend to the complete feedback material.

The fourth variation, in which the response is highlighted in context, was designed on the other hand to allow for a comparison with the findings reported by Krumboltz and Bonawitz (1962). This variation does not differ from the previous one except that the specific response confirmed in context is highlighted by capitalization and underlining.

The fifth variation represented a treatment designed to function as a control procedure. The teaching machine, in this case, was used to

---

* An exception to this situation where programmed booklets are used is when the student has to turn the page before his response can be confirmed. This procedure would be somewhat comparable to the confirmation technique first described.

administer frames of the program which were identical in all respects with the basic learning material and format described above, except that confirmation of the correct response was omitted altogether. Subjects merely proceeded to the next frame after making the required responses.

The following summarizes the five methods of providing confirmation used in the study:

1. Response confirmed in isolation and displayed in absence of the frame material (isolation/frame absent).

2. Response confirmed in isolation in the presence of the frame material (isolation/frame present).

3. Response confirmed is context but not highlighted (context/ not highlighted).

4. Response confirmed in context and highlighted by capitalization and underlining of the correct response(s) (context/ highlighted).

5. Response not confirmed (no confirmation).

Test Materials

Four daily posttests were constructed to cover the material studied during each of four instructional sessions. The number of items in the daily tests ranged from 30 to 40, with a combined total of 120. Each test was composed of 4 subtests. In Subtest I subjects were instructed to recall medical word parts. For example, when presented with "tumor" the subject was expected to respond in writing with its equivalent medical word part "oma." Subtest II reversed the task requiring the subject to supply the meaning of a medical word part. In this case, for example, "carcin" was presented as the test item used to elicit the response "cancer."

The next two subtests expanded the task to include complete medical terms. Subtest III test items were definitional phrases, such as "cancerous tumor", which required the subject to recall the medical term, in this instance "carcinoma." In Subtest IV medical term test items were presented to measure proficiency in the recall of definitions.

A comprehensive post-program test, containing five subtests, was also administered. The first four subtests were designed in the same way as the daily tests. Subtest V was assigned a generalization or transfer function. In this subtest, subjects had to provide definitions for medical terms that were not taught in the program. For example, one of the items included in the test was "melanocarcinoma" which was a unique medical term as far as each subject's experience with the program was concerned. The individual word parts in this term, on the other hand, had been encountered by themselves and as parts of a variety of different medical terms in the program. Subjects were expected to

respond in this instance with "black cancerous tumor", or any approximate definition, indicating their ability to apply the information they had acquired in the program to new situations.

Subtests I and II each comprised 20 items, while Subtests III and IV each comprised 32 items. There were 34 items in Subtest V.

Subjects

The data to be reported were obtained from 96 engineering students enrolled as freshmen at Northeastern University. They were randomly selected from a volunteer subject pool and were paid for their participation. Seventeen females were included among the subjects. Students reporting previous experience in medical or medically allied fields were not allowed to serve as subjects.

Subjects were randomly assigned to one of five treatment groups, with each group having a planned sample size of 20. Five students failed to complete the prescribed sessions. Two of these were students who had been designated for the no confirmation group but did not attend the first session. Three other students dropped out during the instructional sessions for various reasons (2 from the no confirmation group and 1 from the isolation/frame absent group). An administrative mishap increased the size of the isolation/frame present group to 21.

Procedure

Subjects were allowed to schedule themselves for anytime during a 14 hour day, and were required to complete the program in four consecutive daily (Monday through Thursday) sessions. The maximum number of subjects permitted to be scheduled during any single hour was 10. Each subject was given a prescribed number of frames (approximately 95) to complete during each session. The daily tests were given immediately after the completion of the programmed units. The comprehensive post-program test as administered on the fifth consecutive day (Friday).

Each session was scheduled for two hours, but the majority of the subjects were able to complete the assigned frames and daily test in approximately 1½ hours. Subjects reported to a study center where two, and sometimes three, monitors were in constant attendance. On the first day the monitors spent about 10 minutes at the beginning of the session giving individualized instructions to each subject on the use of the device and on the requirements of the program and daily tests. All subjects were instructed to print their responses. Monitors handed out programs and test materials as they were needed and recorded the daily completion times for each subject. On the fifth day monitors again spent a few minutes with each subject to discuss the administrative aspects of the comprehensive post-program test.

The teaching machine used for program presentation was also used for test administration. The device allowed test items, enclosed in frames similar to the programmed material, to be exposed one at a time,

and prevented subjects from seeing items more than once. For both the daily and comprehensive tests, the subtest administration sequence was I, III, II, IV for half of the subjects in each group, and II, IV, I, III for the other half in each group. Subtest V was presented as the last part of the comprehensive test for all groups of subjects. Each subject was instructed to formulate words and definitions to the best of his ability and to spell his test answers as accurately as possible.

Results

Program Performance. Both the amount of time taken by a subject to complete each of the four programmed units and the total number of correct responses he made for each of the units were examined to determine whether program performance was influenced by any of the confirmation variations. The first column of Table II-1 contains the mean number of correct program responses for each of the experimental groups. Misspelled medical terms and inappropriately worded definitions, as well as incorrect or incomplete responses and omissions, were considered as errors. A one-way analysis of variance on the correct response data indicated that the differences across confirmation treatments were not significant, $F(4,91) = 1.18$, $p>.05$. A similar analysis was performed on program completion times. The differences among the means presented in the second column of Table II-1 also were not found to be significant ($F<1$).

TABLE II-1

Program Performance: Mean Number of Correct Responses and Mean Completion Times (Minutes) for the Five Confirmation Procedures

Confirmation Procedure	Correct Responses	Completion Time
Isolation/frame absent	1031.0	328.9
Isolation/frame present	1025.3	309.6
Context/not highlighted	1011.0	332.7
Context/highlighted	1002.3	338.1
No confirmation	992.9	332.0

Test Performance. The response accuracy categories established for the evaluation of the Study 1 results were also used in the present study for scoring test responses. In brief, responses to test items eliciting the recall of medical terms were categorized along a continuum ranging

from correctly spelled terms to responses with spelling inaccuracies involving two letters. The "accurately worded" and "acceptable approximation" categories were used to assess the extent to which definition test responses followed the wording provided by the program. As in the first study, three judges were used to score the tests, and any response falling outside of the prescribed categories was scored as incorrect.

The results of the daily tests and the post-program comprehensive test, except for Subtest V, are presented in Tables II-2 and II-3, respectively. The mean scores are cumulated from left to right across the three categories concerned with medical term responses and the two definition response categories. The four daily test scores were combined to represent an overall measure of immediate retention. In addition, scores from Subtest I (medical word parts) were combined with those of Subtest III (complete medical terms) to provide a mean score for medical terminology test responses. Further, scores from Subtests II (meaning of medical word parts) and IV (meaning of medical terms) were combined for the definition test items. These combined subtests were found to be significantly intercorrelated for all treatment groups in both the daily and post-program tests. The correlation coefficients ranged from +.78 to +.91.

As can be seen from Tables II-2 and II-3, the immediate tests and the post-program test produced essentially the same results. While four of the confirmation treatments appear to have little differential effect on test performance, the mean scores for the context/highlighted group are noticeably lower on both tests for the two types of retention test items. Lower scores for this group are consistent across all of the response accuracy categories. One-way analyses of variance for each of the categories, however, failed to demonstrate any significant differences among any of the confirmation treatments at the .05 level. The results of the statistical analyses are presented in Tables II-4 and II-5.

As stated earlier, Subtest V was designed to measure a subject's ability to define new medical terms composed of combinations of word parts that were practiced in the program. The results of this subtest, considered to reflect proficiency in transfering or applying previously acquired information, are shown in Table II-6. The differentiation among the confirmation treatments closely parallels the previously presented test findings. Again, analyses of variance indicated that the different treatments had no significant effects either when the "accurately worded" definition test responses were considered independently, $F(4,91) = 0.78$, $p > .05$, or when they were combined with the "acceptable approximation" responses, $F(4,91) = 0.76$, $p > .05$.

Discussion

Overall, the results of this investigation are consistent with the findings reported by Holland (1960), Feldhusen and Birt (1962), Krumboltz and Weisman (1962), Hough and Revsin (1963), Ripple (1963), and Rosenstock, Moore and Smith (1965). Significant differences in retention test scores were not found to be associated with the presence versus the

TABLE II-2

Combined Daily Tests: Mean Number of Responses for the Various Response Accuracy Categories as a Function of Confirmation Procedure and Type of Retention Test Item

Confirmation Procedure	Medical Term Test Items			Definition Test Items	
	Accurately Reproduced	0 to 1-Letter Spelling Inaccuracy	0 to 2-Letter Spelling Inaccuracy	Accurately Worded	Accurately Worded and Acceptable Approximation
Isolation/frame absent	52.0	56.8	57.9	58.7	59.2
Isolation/frame present	52.0	57.6	58.2	59.6	60.3
Context/not highlighted	51.0	56.7	57.9	60.4	61.1
Context/highlighted	46.0	51.0	52.9	52.7	53.5
No Confirmation	51.7	58.2	59.7	60.1	61.0

TABLE II-3

Comprehensive Post-Program Test: Mean Number of Responses for the Various Response Accuracy Categories as a Function of Confirmation Procedure and Type of Retention Test Item

Confirmation Procedure	Medical Term Test Items			Definition Test Items	
	Accurately Reproduced	0 to 1-Letter Spelling Inaccuracy	0 to 2-Letter Spelling Inaccuracy	Accurately Worded	Accurately Worded and Acceptable Approximation
Isolation/frame absent	25.5	30.1	31.6	36.8	37.2
Isolation/frame present	25.9	30.4	30.8	35.2	35.8
Context/not highlighted	24.7	29.8	31.1	36.7	37.3
Contest/highlighted	21.3	24.8	26.4	31.7	32.3
No Confirmation	24.8	29.5	30.8	36.9	37.5

TABLE II-4

Summary of Analyses of Variance of the Combined Daily Test Scores

Source	df	Medical Term Test Items						Definition Test Items			
		Accurately Reproduced		0 to 1-Letter Spelling Inaccuracy		0 to 2-Letter Spelling Inaccuracy		Accurately Worded		Accurately Worded and Acceptable Approximation	
		MS	F	MS	F	MS	F	MS	F	MS	F
Between	4	129.99	1.11	150.32	1.51	126.99	1.44	200.89	2.26	194.25	2.32
Within	91	116.61		99.09		88.16		88.55		83.65	

TABLE II-5

Summary of Analyses of Variance of the Comprehensive Post-Program Test Scores

Source	df	Medical Term Test Items						Definition Test Items			
		Accurately Reproduced		0 to 1-Letter Spelling Inaccuracy		0 to 2-Letter Spelling Inaccuracy		Accurately Worded		Accurately Worded and Acceptable Approximation	
		MS	F	MS	F	MS	F	MS	F	MS	F
Between	4	67.00	1.01	108.75	1.68	89.37	1.41	95.70	1.50	93.48	1.50
Within	91	66.18		64.54		63.17		63.59		62.21	

TABLE II-6

Comprehensive Post-Program Test, Subtest V: Mean Number of Definition Test Responses for Two Response Accuracy Categories as a Function of Confirmation Procedure

Confirmation Procedure	Response Accuracy Categories	
	Accurately Worded	Accurately Worded and Acceptable Approximation
Isolation/frame absent	21.1	21.8
Isolation/frame present	20.8	21.6
Context/not highlighted	20.4	21.2
Context/highlighted	17.6	18.4
No Confirmation	20.8	21.4

absence of confirmation during program presentation. The complexity of the responses required by the program does not appear to be an important determinant. The lack of any significant effect was evidenced for the two distinct types of responses: the novel medical terms and the more meaningful definitions. An additional consideration, the ability to reproduce such responses either with absolute accuracy or with varying degrees of preciseness, was not found to be influenced by providing or withholding confirmation.

The contention proposed earlier by the present investigators that previous studies in the area of confirmation might have had their effects nullified through isolating the confirmed response from its appropriate context cannot be supported by the present findings. Retention test scores were not improved either by supplying frame content along with the feedback material or by keeping the frame in view while confirming responses. Further, and contrary to the finding reported by Krumboltz and Bonawitz (1962), providing confirmation in context did not result in superior test performance on test items measuring proficiency in the application of principles taught by the program.

One of the most consistent positive findings from studies investigating the effects of confirmation in program instruction is the significantly poorer performance associated with non-confirmation when program errors are considered, (Krumboltz and Weisman, 1962; Ripple, 1963; Rosenstock, Moore and Smith, 1965; and Jacobs and Kulkarni, 1966). Notably, in all of these studies the programmed subject matter was presented in booklet form. The failure to obtain similar results in the present study raises doubts concerning the adequacy of the research findings cited above while suggesting the possible source of variance in their treatments. Valid error rate comparisons in studies using program booklets are dependent upon the extent to which subjects actually respond before looking at the confirmation material. According to Jacobs and Kulkarni (1966), subjects tend to neglect or ignore the instructions regarding the prescribed use of the confirmation item. These investigators were able to produce evidence that a number of subjects, all of whom used booklets in their study, looked ahead before recording their answers. Unlike programmed booklets, teaching machines of the type used in the present study prevent subjects from observing the confirmation material prior to making responses. The present study, contrary to previous findings, does not lend support to the interpretation that omission of confirmation has a significant effect on program error rate.

IV. The Interactive Effects of Frame Content, Response and Confirmation Variations in Programmed Instruction (Study 3).

Introduction

A persistent research question in programmed instruction since its inception has been the effectiveness of the written response. Is post-program retention significantly enhanced when the learner is required to write his responses instead of merely "thinking" about them? The investigations cited in the overall Introduction to this report revealed this question to be enveloped in a shroud of contradictory results. Of the many attempts to experimentally resolve such findings, not one has been able to account for the inconsistencies.

Reviewers who have attempted to conceptually isolate the variable or variables responsible for the discrepant findings have also failed to meet with any success (cf. Holland, 1965; Lumsdaine and May, 1965; May, 1966; and Anderson, 1967). Even a cursory review of the literature points to the divergencies that characterized the many studies in this area. In these earlier studies variations among such factors as programmed subject matter, program length, population samples, frame design, types of retention tests, etc., have only served to mitigate chances for providing valid resolutions.

The findings of Study 1 of this report as well as those of Eigen and Margolies (1963) and Williams (1965) are consistent, however, in providing support for the hypothesis that overt responding is demonstrably superior to other response modes only when retention tests require the written recall of difficult or technical material learned in the program. It would appear from these results that the question of the efficacy of the written response in linear self-instructional programs may be resolved in terms of response difficulty as the critical variable. That other variables may be responsible, however, has been suggested by a number of investigations.

Cummings and Goldstein (1964), for example, were able to demonstrate the superiority of overt over covert responding with a program on the diagnosis of myocardial infarction. These investigators suggested that the observable record provided by overt responding was primarily responsible for the superior results. They felt that responding by writing enabled the learners to directly compare their responses with the correct ones supplied by the confirming information, thereby facilitating the learning of the difficult verbal and pictorial (EKG) responses required by the program. Study 2 in the present report, however, did not provide any evidence of a significant interaction involving response mode, response difficulty and confirmation procedure to substatiate the Cummings and Goldstein interpretation. Confirming technical terms in a program which required the learner to write his responses did not result in any better retention than when confirmation was withheld.

Kemp and Holland (1966) considered frame content to be one of the most important determinants governing the effect of response mode. Using their frame "blackout ratio" technique (Holland and Kemp, 1965),

they asserted that the effect of overt responding was a function of how relevant the content of the entire frame was to the response being elicited. A number of programs previously administered in overt vs. covert experimental comparisons was used in their study to demonstrate that the more irrelevant the material included in program frames, the less effective the program was in showing the superiority of overt responding on retention test scores. The level of difficulty of the responses elicited by the various programs used in their analysis was not considered.

Finally, as indicated by Krumboltz and Weisman (1962), there is the possibility that a delayed retention test may be more sensitive to the effects of response mode, notwithstanding the effects of other variables. Subjects who were required to respond overtly in their investigation, while exhibiting no significant degree of superior performance for immediate recall, were observed to perform significantly better than subjects who responded covertly on a two-week retention test. The retention test administered in the first study of the present report, which also demonstrated the superiority of overt over covert responding, was given one week after program completion. The determination of whether response mode and response difficulty interact with delayed retention requires additional study.

In the present investigation a factorial design was used to specify further the relationship between response mode and response difficulty. Three other independent variables: a) frame content, b) number of required responses per frame, and c) confirmation procedure were incorporated into the design to determine their independent effects and to assess their possible interaction with the response mode and difficulty treatments. In addition, a series of retention tests administered during the program and after its completion were used to analyze the effects of these program variations over time.

Two levels of each independent variable were investigated. The response mode treatment called for either overt or covert responding. Response difficulty was dichotomized by utilizing a program which required two distinguishable types of responses: a) technical medical terms, and b) nontechnical definitions of medical terms. Different versions of the program were developed to manipulate frame content. In one version each frame was restricted to the wording necessary for response elicitation, while in the other, additional but irrelevant material was added to each frame. The programs were further subdivided into versions that required one response per frame and others in which more than one response was necessary. The confirmation treatment variation was minipulated by either providing or withholding information on the correct responses. In addition, reading programs, which were not included as an integral part of the factorial design, were used to provide comparative control data.

Subjects

Subjects were drawn from a pool of students who had indicated a willingness to participate in psychological experiments in response to a questionnaire. All were incoming freshmen at Northeastern University, Boston. Foreign students and students who reported that they were going

into premedical or medically allied programs were withdrawn from the pool. The remaining students were enrolled in a wide variety of science and non-science programs.

Subjects were randomly assigned on a weekly basis to one of 18 groups until each group had reached a size of 25. Additional students had to be selected from the pool to replace 41 subjects who, for reasons to be explained later, failed to complete all of the prescribed experimental sessions. The final sample was made up of 253 males and 197 females. Each had been questioned prior to the experiment and none reported anything other than a layman's knowledge of medical terminology. All subjects were paid for their services.

Instructional Materials

A revised version of the linear program on medical terminology employed in Study 2 was used in the present investigation. Since all of the treatments in the second study involved written program responses, it was possible to use the error rate data of approximately 100 subjects as the basis for the revision. All frames with an error rate greater than 5% were either rewritten or supported by additional frames, or in some cases deleted altogether. As a result of this modification, the program consisted of 384 frames in which 148 medical words were developed from 62 Greek and Latin word parts.

The techniques employed to teach the construction and meaning of two of the medical terms in the program can be used to illustrate the basic teaching paradigm followed in devising the instructional sequences. In teaching the medical term acromegaly, acro is introduced first as a combining form used in medical terms to refer to the bodily extremities. The frames dealing with this word part are followed by a second set of frames discussing another medical word part, megal, presented as the word root meaning that something is enlarged. The student is then given the information that the suffix y can be used as a noun ending. Finally, the complete medical term acromegaly is developed. After practicing with this term and its meaning, the student is introduced to the next sequence involving dermat and itis, respectively, as the word parts found in medical terms referring to the skin and an inflamed condition. The word part acro is then brought back for review, facilitating the development of the medical term acrodermatitis. Once a particular frame sequence has been developed, its medical word parts are incorporated into succeeding sets of frames where they are used to form different medical terms. For example, the combining form acro, which is first introduced in frame 33, is used to construct the terms acromegaly (frames 39, 43-45, 53 and 55), acrodermatitis (frames 48, 54-57, and 70), and acrocyanosis (frames 60-66, and 72). The word root megal, which first appears in frame 37, is used in conjunction with acro to form the term acromegaly, and after being reviewed is used to form the terms megalocardia (frames 90-101) and megalogastria (frames 98-101). The word root dermat, which initially appears in frame 46, is later incorporated into the words acrodermatitis, dermatitis, (frames 67-70) and dermatosis (frames 71, 72, 74 and 75). The suffix itis is first introduced as part of the term acrodermatitis in frame 48 and is then used with other word parts throughout the program to

form the words adenitis, arthritis, cheilitis, cystitis, cholecystitis, dermatitis, encephalitis, gastritis, gingivitis, gingivoglossitis, glossitis, laryngitis, osteochondritis, otitis, pharyngitis, rhinitis, and stomatitis. Thus, word parts are interspersed throughout the program to be used as cumulative components to form medical words. In some frame sequences the medical words are formed either during or after the presentation of their constituent word parts. Other sequences require the student to recall or review previously learned word parts while new information is being presented in the remaining components of a newly introduced medical term.

The basic program served as the prototype for the development of the six experimental programs described below.

1. Multiple Response, Basic Frame Program

The essential experimental features of this program were: (1) it contained relatively short frame lengths, and (2) the vast majority of the frames required more than one response. A frequency distribution of the number of frames per frame-size interval is shown on the left side of Fig. 1. A percentage distribution of the number of responses required per frame is presented in Table III-1.

TABLE III-1

Percentage of Frames Requiring From 1 to 5 Responses in the Multiple Response, Basic Frame Program (384 frames)

Number of Responses Required Per Frame	Percent of Frames
1	1
2	27
3	46
4	21
5	5

A total of 1,154 responses were required by the program. Of these, 618 were concerned with either medical terms or medical word parts, and 434 with complete or partial definitions of medical terms. The remaining 102 responses, hereafter referred to as instructional term responses, dealt with the exposition of the rules governing the use of prefixes, suffixes, word roots and combining forms in building compound words. The majority of these responses were required in the first 29 frames, which were used to provide a review of word building principles employing common English words as examples.

The Multiple Response, Basic Frame Program was subjected to the following four experimental treatments:

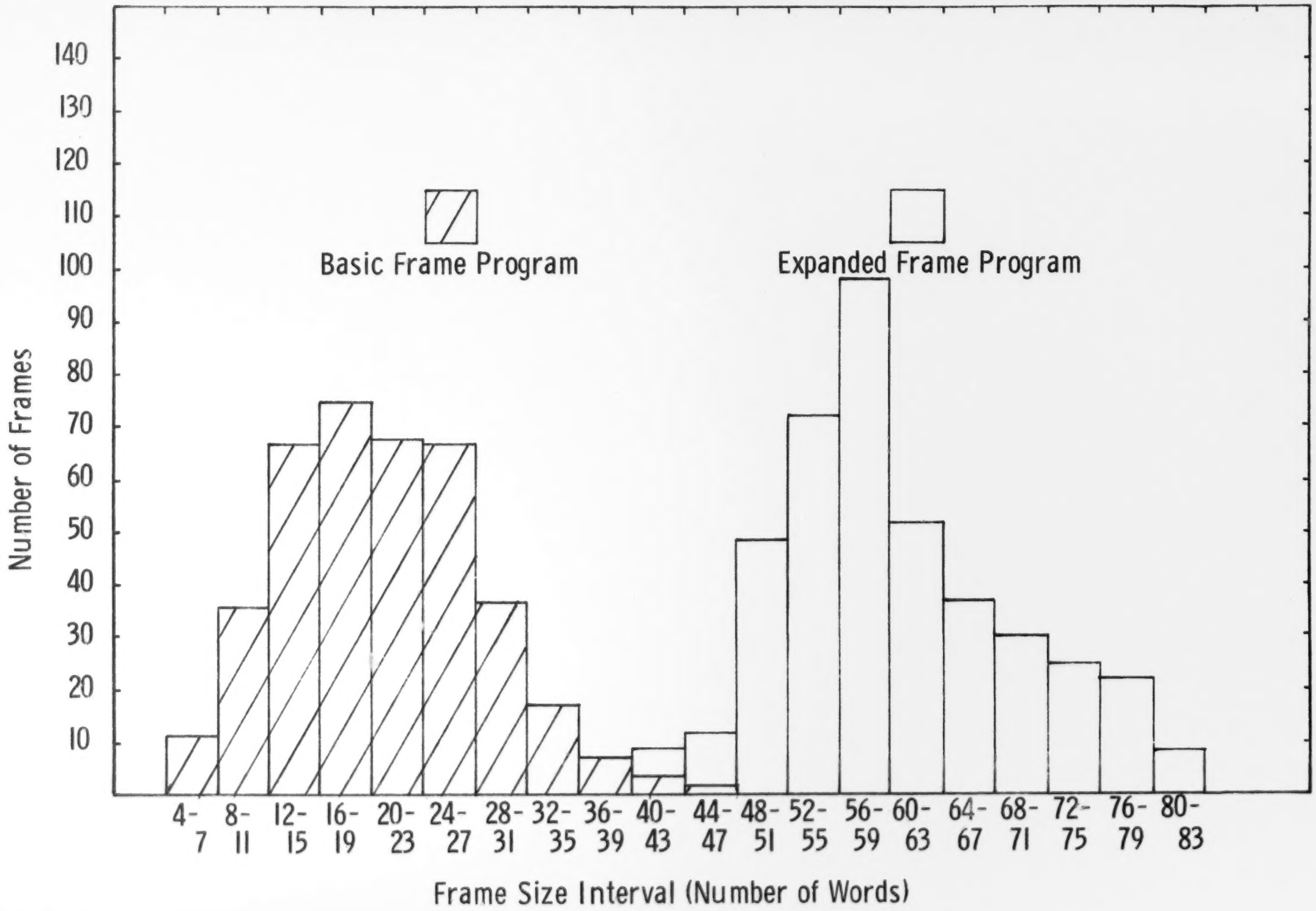

Fig. 1. Distribution of Frame Lengths for the Basic Frame and Expanded Frame Programs

a. Overt response - confirmation. The subjects in this group were instructed to complete the frame material by making written responses in a space provided below the instructional frame. They were allowed to compare their responses with the correct ones which appeared after the instructional frame was manually advanced. A three-frame sequence of this program is presented in Fig. 2.

b. Overt response - no confirmation. This program also required an overt response, but a black over-print covering the confirmation material prevented subjects from obtaining information regarding response correctness.

c. Covert response - confirmation. The subjects who were instructed by this program were told to think about the responses that would best fill in the blanks appearing in the frame material. They were provided with the program used by the overt response - confirmation subjects which allowed them to confirm their responses.

d. Covert response - no confirmation. The subjects in this group were given the same program as those in the overt response - no confirmation group. However, instead of writing their responses, these subjects were instructed only to think of the correct responses.

2. Single Response, Basic Frame Program

This version was designed to be a short frame-length program requiring one response per frame. Essentially, the material was identical to the Multiple Response, Basic Frame Program. However, only one response was left blank for the subject to complete, the remainder in each frame having been filled in with the appropriate information.

A second departure from the Multiple Response Frame Program became necessary. Filling in all but one of the response blanks in many of the frames designed to serve as criterion or unprompted recall frames created too many cues for the correct response. Consequently, holding to a one-to-one correspondence of frames between Multiple and Single Response Frame Programs only served to curtail the actual number of criterion frames in the latter program. To provide for criterion performance and still avoid the overprompting feature of the program, such frames in the Single Response Frame Program were constructed by dividing the corresponding material in the multiple response frames into two parts. The result was two consecutive frames from these parts, each requiring one response. Thirty-eight frames from the Multiple Response Frame Program were treated in this manner. Consequently, the Single Response, Basic Frame Program contained a total of 422 frames. The response requirement in 259 of the frames involved the construction of medical terms. Responses dealing with definitions of medical terms were elicited by 125 frames, and 38 frames required instructional responses. Each medical term and each definition included in the program was required as a response in at least one frame.

Subjects instructed by the Single Response, Basic Frame Program were also divided into four groups, namely, overt response - confirmation, overt response - no confirmation, covert response - confirmation and

198. abdomin/o is used to build words about the abdomen. When you see abdomin or _____/__ any place in a word, you should think about the _____.	
1. abdomin/o 2. abdomen	1. 2.
PROCEED TO NEXT FRAME	
199. An abdomin/o/centesis is a surgical puncture of the _____. Since _____/____ refers to the abdomen,_____ must mean surgical puncture.	
1. abdomen 2. abdomin/o 3. centesis	1. 2. 3.
PROCEED TO NEXT FRAME	
200. Centesis (surgical puncture)is a word in itself. The medical word for surgical_____ of the abdomen, namely abd _ _ i _/o/centesis, is made up of the combining form _____and the word _____.	
1. puncture 2. abdomin/o/ centesis 3. abdomin/o 4. centesis	1. 2. 3. 4.
PROCEED TO NEXT FRAME	

Fig. 2. Frame sequence from the Multiple Response, Basic Frame Program.

covert response - no confirmation. A three-frame sequence from the overt - confirmation version which corresponds to the sequence presented for the Multiple Response Frame Program is shown in Fig. 3.

3. Basic Frame Reading Program

This version was essentially a sequential, short frame reading program. It was identical to the 422-frame Single Response, Basic Frame Program except that the response blank in each frame was already filled in. This meant that no response was required and consequently, no need for confirmation material. An example of some of the reading frames is presented in Fig. 4.

4. Multiple Response, Expanded Frame Program

The previously described Basic Frame Programs were carefully constructed to insure that each frame was restricted to material directly related to the elicited responses. The result was, in each case, a program with relatively short frames: from 4 to 48 words in length. The Multiple Response, Expanded Frame Program was developed by treating the content in each frame of the Multiple Response, Basic Frame Program to expansion techniques: declarative circumlocution, superfluous description, and the infusion of additional but inconsequential information into each frame. Expansion was governed by the constraint to keep the Basic Frame Program frames relatively intact in syntax and wording, adding material to frames only in ways which precluded any additional assistance in eliciting the required responses. The method used in the majority of the frames was to provide information about the etiology, characterisitic symptoms, examination procedures or the prescribed treatments involved in the various afflictions represented by the medical terms in the program.

Developing this kind of material necessitated strict adherence to a set of rules and guidelines. For example, in constructing these frames any additional material had to avoid direct association with the responses being elicited. The information in the contextual addition could not be allowed in any way to represent a cue or prompt, or to repeat any material relevant to eliciting a response, nor could it be allowed to provide the subject with extra instruction on the meaning and construction of medical terms. When words such as "pain" were part of the relevant material in a particular Basic Program frame, the material introduced to create the expanded frame version used neutral, non-cueing synonyms such as "feeling" and noncommittal pronouns such as "it". Medical terms, whether or not contained in the program, could not be included as part of the added frame content. Further, to prevent any kind of covert responding, the additional material was never permitted to be interrogative, or to direct the subject's recall to previously exposed information. Finally, all of this extra material had to be maintained at a level that would keep the subjects interested in reading the frames carefully.

The frames in this program ranged from 40 to 83 words in content. The frequency distribution of frame sizes is presented on the right

198.	abdomin/o is used to build words about the abdomen. When you see abdomin or abdomin/o any place in a word, you should think about the _____.
abdomen	_____
PROCEED TO NEXT FRAME	
199.	An abdomin/o/centesis is a surgical puncture of the abdomen. Since _____/___ refers to the abdomen, centesis must mean surgical puncture.
abdomin/o	_____
PROCEED TO NEXT FRAME	
200.	Centesis (surgical puncture) is a word in itself. The medical word for surgical puncture of the abdomen, namely abdomin/o/centesis, is made up of the combining form abdomin/o and the word _____.
centesis	_____
PROCEED TO NEXT FRAME	

Fig. 3. Frame sequence from the Single Response, Basic Frame Program.

198.	abdomin/o is used to build words about the abdomen. When you see abdomin or abdomin/o any place in a word, you should think about the abdomen.
	PROCEED TO THE NEXT FRAME
199.	An abdomin/o/centesis is a surgical puncture of the abdomen. Since abdomin/o refers to the abdomen, centesis must mean surgical puncture.
	PROCEED TO THE NEXT FRAME
200.	Centesis (surgical puncture) is a word in itself. The medical word for surgical puncture of the abdomen, namely abdomin/o/centesis, is made up of the combining form abdomin/o and the word centesis.
	PROCEED TO THE NEXT FRAME

Fig. 4. Frame sequence from the Basic Frame Reading Program.

side of Fig. 1. It can be noted that the difference in frame size between the Basic and Expanded Frame Programs is appreciable. As previously indicated, the two programs are identical in the number of frames comprising the program and in the specific responses required by each frame.

The examples in Fig. 5 present expanded frames that can be compared with their smaller counterparts in Fig. 2. The Multiple Response, Expanded Frame Program also generated four experimental treatments involving two response modes and the presence and absence of confirmation.

5. Single Response, Expanded Frame Program

This program was devised to represent a large frame program requiring one response per frame. It was identical to the Multiple Response, Expanded Frame Program except that, as in the Single Response, Basic Frame Program, all but one of the response blanks were filled in, and the number of frames was increased to 422. The specific responses required by each frame were identical in the Single Response, Basic and Expanded Frame Programs.

An illustration of frames from this program is provided in Fig. 6. The experimental manipulation of response mode and confirmation also created four treatment conditions for the Single Response, Expanded Frame Program.

6. Expanded Frame Reading Program

This program was identical to the Multiple Response, Expanded Frame Program except that, as can be seen from the examples provided in Fig. 7, all of the response blanks were filled in with the appropriate responses. This version represented a 384 large frame reading program.

The programs were commercially reproduced on 8½ x 11 inch paper by an offset printing process. Each program page contained three frames. A TMI-Grolier, Min/Max III Teaching Machine was used to administer each version of the program.

Test Materials

Each subject was tested for proficiency in the recall and written reproduction of the programmed subject matter at three different time periods. The first series of retention tests were administered as posttests immediately after the completion of each program unit, with each posttest being divided into a number of subtests. The items in one of the subtests were used to measure proficiency in the recall of medical terms, given the definitions of these terms. In the second subtest the items were concerned with the recall of definitions, given the medical terms.

198. abdomin/o is used to build words about the abdomen. This is the cavity of the body that lies below the chest and above the pelvis. It contains many of the vital organs of the body. When you see abdomin or _____/_____ any place in a word, you should think about the _____.	
1. abdomin/o 2. abdomen	1. 2.
PROCEED TO NEXT FRAME	
199. When a physician examines this area during a physical examination, he feels for areas of tenderness or rigidity, evidence of fluid, and abnormal elevations or depressions. If certain conditions prevail, then the patient may require special medical treatment. An abdomin/o/centesis is a surgical puncture of the _____. Since _____/_____ refers to the abdomen, _____must mean surgical puncture.	
1. abdomen 2. abdomin/o 3. centesis	1. 2. 3.
PROCEED TO NEXT FRAME	
200. If during a physical examination a physician finds that an area is very tender so that the patient cringes when it is touched, the physician may require that the patient undergo further tests to determine the cause of the ailment. He may require the special treatment that was mentioned in the previous frame. Centesis (surgical puncture) is a word in itself. The medical term for surgical _____ of the abdomen, namely abd _ _ i _/o/-centesis, is made up of the combining form _____/_____ and the word _____.	
1. puncture 2. abdomin/o/centesis 3. abdomin/o 4. centesis	1. 2. 3. 4.

Fig. 5. Frame sequence from the Multiple Response, Expanded Frame Program.

198. abdomin/o is used to build words about the abdomen. This is the cavity of the body that lies below the chest and above the pelvis. It contains many of the vital organs of the body. When you see abdomin or abdomin/o any place in a word, you should think about the _____.	
abdomen	_____
PROCEED TO NEXT FRAME	
199. When a physician examines this area during a physical examination, he feels for areas of tenderness or rigidity, evidence of fluid, and abnormal elevations or depressions. If certain conditions prevail, then the patient may require special medical treatment. An abdomin/o/centesis is a surgical puncture of the abdomen. Since _____/ refers to the abdomen, centesis must mean surgical puncture.	
abdomin/o	_____
PROCEED TO NEXT FRAME	
200. If during a physical examination a physician finds that an area is very tender so that the patient cringes when it is touched, the physician may require that the patient undergo further tests to determine the cause of the ailment. He may require the special treatment that was mentioned in the previous frame. Centesis (surgical puncture) is a word in itself. The medical term for surgical puncture of the abdomen, namely abdomin/o/centesis, is made up of the combining form abdomin/o and the word _____.	
centesis	_____

Fig. 6. Frame sequence from the Single Response, Expanded Frame Program.

198. abdomin/o is used to build words about the abdomen. This is the cavity of the body that lies below the chest and above the pelvis. It contains many of the vital organs of the body. When you see abdomin or abdomin/o any place in a word, you should think about the abdomen.

PROCEED TO NEXT FRAME

199. When a physician examines this area during a physical examination, he feels for areas of tenderness or rigidity, evidence of fluid, and abnormal elevations or depressions. If certain conditions prevail, then the patient may require special medical treatment. An abdomin/o/centesis is a surgical puncture of the abdomen. Since abdomin/o refers to the abdomen, centesis must mean surgical puncture.

PROCEED TO NEXT FRAME

200. If during a physical examination a physician finds that an area is very tender so that the patient cringes when it is touched, the physician may require that the patient undergo further tests to determine the cause of the ailment. He may require the special treatment that was mentioned in the previous frame. Centesis (surgical puncture) is a word in itself. The medical term for surgical puncture of the abdomen, namely abdomin/o/centesis is made up of the combining form abdomin/o and the word centesis.

Fig. 7. Frame sequence from the Expanded Frame Reading Program.

The posttests following the first three of four consecutive daily units of program administration included an additional subtest that was not designed to produce data. It consisted of a small number of short essay items such as "What information did the medical terminology program provide about cancerous tumors?" and "Why would a surgeon perform an abdominocentesis?". These questions were intended to control for the maintenance of interest in, and attention to, the peripheral material appearing in the expanded frame versions of the program. The questions would draw for their answers on the subject matter in all of the programmed versions; the expanded frame versions, however, contained more information that could be utilized in the answers. It was felt that in the absence of this type of test item, subjects taught by the Expanded Frame Programs would tend to adopt strategies to circumvent studying the additional information and attend only to those aspects of the frames concerned with either medical terms or their definitions.

The items comprising each of the four posttests were restricted to the material covered in the programmed unit administered on a particular day. The essay section was not included as part of the posttest on the fourth day of program presentation.

All of the test items requiring medical terms and medical term definitions in the daily tests were selected to represent informationally independent items. For example, in each posttest the medical terms selected as subtest items were constructed in every case by dissimilar word parts. None of these word parts were repeated in the subtest requiring definitions as responses. For half of the subjects in each experimental group the subtest containing the medical term definition items was given first, followed by the subtest requiring the recall of medical terms. The reverse order was utilized for the remaining subjects. Each posttest contained six medical term items and six medical term definition items, except for the posttest administered after the first programmed unit which was composed of five medical terms and five definition items. The essay section of the posttests was always the last administered on each day that it was included.

On the day immediately following the fourth programmed instructional unit a comprehensive retention test which sampled the material covered in the entire program was administered. This test contained 56 items, 95% of which were different from the items included in the daily posttests, and was presented in four parts. Part I (18 items) and Part III (10 items) presented medical terms, and subjects were instructed to provide their definitions. Part II (18 items) and Part IV (10 items) presented definitions and required the recall of medical terms. Items in Part I and Part II were selected to be informationally independent; that is, there was minimal correspondence between the medical terms of one part and the definitions of the other part. Parts III and IV were included primarily to increase the number of items in the test. Since only 62 different word parts were taught in the program, the items in Parts III and IV were not informationally independent of those in the first two parts. For half of the subjects in each experimental group the order of test presentation was Parts I, II, III, IV; while the sequence II, I, IV, III was utilized for the remainder of the subjects.

Subjects were scheduled to return from 32 to 34 days after the comprehensive retention test to take a long term retention test. This test was also designed to include items from all of the four units of the program. It was composed of 50 items, 82% of which were different from the items that appeared in the four posttests and the comprehensive retention test. The long term retention test was constructed with the same considerations that applied to the comprehensive test, and likewise it was presented in four parts. Parts I and III, which contained 15 and 10 items respectively, required medical term definitions as responses. Part II with 15 items, and Part IV with 10 items required medical term responses. As with the comprehensive retention test, half of the subjects in each group received different test part sequences.

To control for the possibility that differences in medical term and definition test scores could be attributable to intrinsic differences in difficulty in the subtest items themselves, the investigators constructed two different sets of items for each of the retention tests. The definition items in subtests II and IV, which were used to elicit medical term responses for half of the subjects in each experimental group, were presented as the corresponding medical terms and were used to elicit definition responses in subtests I and III for the other half of the group. Conversely, the medical terms appearing in subtests I and III for the former group of subjects corresponded to the definitions in subtests II and IV for the latter group.

In addition to the considerations indicated above, other precautions were taken to minimize the possibility of subjects being aided in their recall of specific test items through exposure to other items in the test. The tests were administered in the same teaching machine which prevented subjects from going back to previously exposed items during program presentation. Further, the test material was constructed so that only one item could be exposed at a time.

Each subject was instructed to print and spell all of his test answers to the best of his ability, and to use, as accurately as possible, the wording of the program as a source for his answers to the examination items.

Procedure

Subjects were required to participate over a period of five consecutive days and to return approximately five weeks later for the delayed retention test. The instructional sessions were conducted during the first four days, followed by the comprehensive retention test on the fifth day.

Prior to participating in the study each subject was randomly assigned to one of the 18 program variations, and scheduled for each of the daily sessions. The programs and tests were administered in a study center where monitors were in constant attendance in a supervisory and surveillance capacity. The number of subjects scheduled during any particular hour was restricted to ten.

On the first day subjects were given instructions, individually, concerning the operation of the teaching machine, and were familiarized with the characteristics of the program and its requirements by going through the first three frames under the guidance of a monitor. Subjects were then given frames 4 through 101 on the first day, followed by frames 102-197, 198-281 and 282-384 on the second, third and fourth day, respectively. Although the 422 frames in the Basic Frame Reading Program and the Single Response Frame Programs were numbered differently, the subjects instructed by these programs received equivalent amounts of material in each daily session.

Subjects assigned to the overt response groups were instructed to print their responses. Those in the covert response and reading groups were not allowed to do any writing while taking the program. All subjects were told that they could work at their own pace. A record was kept of the time subjects took to complete each programmed unit and each retention test.

Each subject was given a portion of his fee after the completion of the five day session. The amount withheld was paid when the subject returned for the long term retention test.

Results

Subject Loss

As noted earlier, data from 41 subjects could not be included in the final analysis. Twenty-eight subjects did not report for all of the instructional sessions, seven were absent from the comprehensive testing session, and six were disqualified from further participation for not following instructions. Chi-square analyses indicated that neither of the reading programs nor any of the program variations created by manipulating frame size, number of required responses per frame, response mode or confirmation were significantly associated with failure to satisfactorily complete the experimental sessions.

Program Completion Time

Table III-2 presents the mean times in minutes to complete the four programmed units for the program variations in frame size, responses per frame, response mode and confirmation. The results obtained from a 2 x 2 x 2 x 2 analysis of variance of the time data were consistent with expectations based upon the procedures used in developing the different programs. Significant main effects were found for Frame Size, Number of Responses and Response Mode. The Expanded Frame Programs, which were developed by adding material to each frame in the Basic Frame Programs, took longer to complete, $F(1,384) = 158.21$, $p < .001$. Increasing the number of required responses per frame, and requiring written responses also lengthened the time needed for program completion; $F(1,384) = 137.65$, $p < .001$ for Number of Responses, and $F(1,384) = 157.50$, $p < .001$ for Response Mode. In addition, a Response Mode X Number of Responses interaction, $F(1,384) = 49.01$, $p < .001$, indicated that the differences observed between the overt and covert response groups in program completion times became significantly more pronounced as the

TABLE III-2

Means and Standard Deviations of Program Completion Times in Minutes as a Function of Frame Size, Number of Responses Per Frame, Response Mode and Confirmation.

FRAME SIZE	NUMBER OF RESPONSES	RESPONSE MODE	CONFIRMATION	M	SD
BASIC	SINGLE	OVERT	Conf.	171.28	42.22
			No Conf.	164.80	37.32
		COVERT	Conf.	141.04	42.97
			No Conf.	129.40	17.69
	MULTIPLE	OVERT	Conf.	242.80	50.63
			No Conf.	238.84	42.12
		COVERT	Conf.	159.48	30.54
			No Conf.	151.64	44.73
EXPANDED	SINGLE	OVERT	Conf.	222.84	54.68
			No Conf.	194.56	35.38
		COVERT	Conf.	193.92	52.11
			No Conf.	195.68	41.08
	MULTIPLE	OVERT	Conf.	292.52	39.43
			No Conf.	294.80	47.49
		COVERT	Conf.	216.36	39.75
			No Conf.	212.24	41.39

number of responses per frame increased. No other significant sources of variance were revealed by this analysis.

The mean completion times in minutes for the Basic and the Expanded Frame Reading Programs were 145.16 (SD = 62.50) and 183.48 (SD = 45.49), respectively. Using a pooled error term obtained from all 18 groups, Dunn's test for multiple comparisons among means (Dunn, 1961) was used to include the Reading Programs in the comparison of program completion times. A significant difference was found between the Basic and the Expanded Frame Reading Programs ($p < .05$).

The mean program completion time for the Basic Frame Reading Program was found to be significantly shorter ($p < .05$) than the mean times for all of the other programs except for the four Single Response, Basic Frame Programs, and the two Multiple Response, Basic Frame Programs which did not require overt responses. These latter programs did not differ significantly ($p < .05$) from the Basic Frame Reading Program. The Expanded Frame Reading Program was found to have a significantly shorter mean completion time than the four Multiple Response, Expanded Frame Programs, the two Multiple Response, Basic Frame Programs which required overt responses, and the Single Response, Expanded Frame Program which required overt responses and provided confirmation. The Expanded Frame Reading Program did not differ significantly from the following six programs: the Single Response, Expanded Frame Programs, except for the overt response/confirmation version; the Single Response, Basic Frame Programs which required overt responses; and the covert response/confirmation version of the Multiple Response, Basic Frame Program. In all other comparisons the Expanded Frame Reading Program exhibited the significantly longer mean completion time.

Retention Test Performance

A slight modification of the test scoring procedure used in Studies I and II was used to evaluate the retention test results obtained from the present study. In these earlier studies definitions that were essentially correct but deviated from the wording used in the program were scored as "acceptable approximation" responses. This category, however, did not prove to be any more effective than the "accurately worded" category alone in differentiating among the effects of the experimental treatments. Consequently, the "acceptable approximation" designation was not included in the present analysis. Instead, spelling inaccuracies provided the basis for scoring definition responses as well as medical terms.

Two criteria were used to evaluate the definition test responses. First, the words used in defining a medical term had to coincide or be synonymous with the wording used in the program. Second, each word used in the definition had to be accurately spelled. Definitions which met both criteria were categorized as "accurately reproduced" responses and distinguished from definitions which met the first criterion but contained one or more misspelled words.

Medical terms that were correctly recalled and accurately spelled

were also recorded as "accurately reproduced" responses. A term that was misspelled but was still judged to be recognizable as the required response was classified along an accuracy-of-response continuum. If one letter in a medical term was either incorrect, omitted, added or transposed, the response was regarded as a "1-letter spelling inaccuracy". When a term contained more than a "1-letter spelling inaccuracy," each incorrect letter in the term was then regarded as a "2-letter spelling inaccuracy." In other words, the subject was penalized for both adding one letter and omitting the correct letter. Each incorrect letter in the term (as a "2-letter spelling inaccuracy") was then summed up with other letters that were either omitted, added or transposed to determine the total number of letters involved in the misspelling.

Three judges who were unaware of the program variations in the study were used to evaluate the test results. All three judges had to agree that a misspelled medical term was still recognizable as the required response before it was scored to determine the number of letters involved. Similarly, complete unanimity was required in determining the acceptability of a definition which did not utilize the wording provided by the program. If one of the judges failed to concur, the medical term was scored as incorrect. Misspelled medical term responses that were classified as representing more than a 5-letter inaccuracy led to consistently unreliable judgments; consequently any response with more than a 5-letter spelling inaccuracy was scored as incorrect.

Table III-3 shows the results of the "accurately reproduced" definition test responses for the program variations in frame size, number of required responses per frame, response mode and confirmation. Table III-4 contains the same breakdown for the definition responses that were correctly recalled but were not scored with the "accurately reproduced" criterion. The scores from the four unit tests were combined to represent a measure of immediate retention, and are presented along with the comprehensive and delayed test scores in these tables.

The combined daily test results of the subtests eliciting medical terms as responses are presented in Tables III-5 and III-6 for the Basic and Expanded Frame Programs, respectively. The comprehensive test results for the medical term responses appear in Tables III-7 and III-8, while data for the delayed retention test appear in Tables III-9 and III-10.

The procedure used in entering the means into the two response accuracy classifications for the definition responses and the four classifications for the medical term responses allows for a comparison of the results obtained when different criteria are used to determine the adequacy of test item responses. The "accurately reproduced" classification for the definition and medical term responses provides results that are obtained when a stringent criterion is used for test item evaluation. Scores that are entered into this classification indicate that subjects were not only able to recall the appropriate responses, but were able, as well, to produce them without error. For definition responses the means in Table III-4 represent the "accurately reproduced" responses in addition to those containing errors in

spelling. Consequently Table III-4 represents the results obtained when a more lenient criterion is used to score the test items. Similarly, the means in Tables III-5 through III-10 are cumulated across the response accuracy classifications to demonstrate the effect on medical term retention scores of decreased standards for spelling accuracy.

Analyses of variance (2 x 2 x 2 x 2) were used to determine the effects of frame size, number of responses per frame, response mode and confirmation on test performance. Separate analyses were performed on each of the medical term and definition response accuracy classifications for each of the three retention tests.

The results of the analyses of the "accurately reproduced" definition test responses are presented in Table III-11. In all three retention tests, only the main effect of Response Mode for the combined unit tests was observed to be significant. More "accurately reproduced" definition responses were made by the subjects instructed by the overt response programs. However, as can be seen in Table III-12 the superiority of written responding is no longer evidenced when definition test responses are scored without taking spelling errors into consideration

The medical term subtest responses, like the definition results, were analyzed through separate four-way analyses of variance for each response accuracy classification. The results of the analyses of the combined unit test data which appear in Table III-13 show that only the main effects for Number of Responses and Response Mode reached significance. Requiring overt responses and increasing the number of responses per frame resulted in higher retention scores across all response accuracy classifications. None of the interaction effects in these analyses were significant.

An important finding concerning the effects of response modes is revealed in Tables III-5 and III-6. An examination of the means in these tables shows that the response mode groups are clearly separated when comparisons are made among "accurately reproduced" responses. The magnitudes representing differences between the response mode groups, however, do not remain constant across all the response accuracy classifications. It can be seen that as increasingly less insistence is placed upon spelling accuracy, the differences between overt and covert responding become progressively smaller. The effect can be observed for both the single and the multiple response variations of the Basic and the Expanded Frame Programs. Irregularities in two of the comparisons involving non-confirmation programs are the only exceptions. Even when the scoring procedure allowed up to a 5-letter spelling inaccuracy for comparisons between the overt and covert response groups, the smaller differences observed for the main effect of Response Mode remained significant.

The increasing convergence between the response mode group means as response reproduction accuracy decreases can also be noted by an examination of the comprehensive test results which appear in Tables III-7 and III-8. This phenomenon can be consistently observed among all the program variations without exception.

TABLE III-3

Means and Standard Deviations of Accurately Reproduced Definition Test Item Responses for Three Retention Tests as a Function of Frame Size, Number of Responses Per Frame, Response Mode and Confirmation Procedure.

RETENTION TEST	RESPONSE MODE		BASIC FRAME PROGRAMS				EXPANDED FRAME PROGRAMS			
			SINGLE RESPONSE FRAMES		MULTIPLE RESPONSE FRAMES		SINGLE RESPONSE FRAMES		MULTIPLE RESPONSE FRAMES	
			CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
COMBINED UNIT TESTS	OVERT	M	18.0	16.6	18.0	17.5	17.4	17.2	18.2	18.0
		SD	(3.2)	(4.1)	(3.2)	(3.0)	(3.2)	(2.4)	(2.2)	(2.6)
	COVERT	M	16.0	16.9	17.0	17.0	17.0	17.6	16.5	16.5
		SD	(4.6)	(3.0)	(2.3)	(3.3)	(3.1)	(3.5)	(3.4)	(3.8)
COMPRE-HENSIVE TEST	OVERT	M	17.6	16.3	19.0	18.8	17.6	17.2	17.7	17.4
		SD	(5.2)	(6.1)	(4.5)	(4.8)	(5.1)	(4.3)	(3.8)	(3.9)
	COVERT	M	15.7	17.2	17.4	16.5	16.7	18.2	16.9	15.5
		SD	(6.3)	(5.0)	(5.0)	(5.1)	(4.5)	(5.3)	(4.0)	(5.7)
DELAYED TEST	OVERT	M	10.2	8.0	9.7	9.3	9.5	9.2	8.6	8.8
		SD	(4.9)	(5.1)	(4.3)	(4.8)	(5.5)	(4.6)	(4.3)	(4.4)
	COVERT	M	7.3	7.9	8.6	8.4	9.2	9.1	8.4	8.9
		SD	(5.6)	(4.1)	(5.3)	(5.4)	(5.3)	(5.1)	(3.4)	(4.5)

TABLE III-4

Means and Standard Deviations of Correctly Recalled Definition Test Item Responses Scored Without Regard for Spelling Accuracy for Three Retention Tests as a Function of Frame Size, Number of Responses Per Frame, Response Mode, and Confirmation Procedure.

RETENTION TEST	RESPONSE MODE		BASIC FRAME PROGRAMS				EXPANDED FRAME PROGRAMS			
			SINGLE RESPONSE FRAMES		MULTIPLE RESPONSE FRAMES		SINGLE RESPONSE FRAMES		MULTIPLE RESPONSE FRAMES	
			CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
COMBINED UNIT TESTS	OVERT	M	19.1	17.8	19.5	19.2	18.9	18.4	19.4	19.5
		SD	(3.4)	(3.9)	(2.2)	(2.9)	(2.9)	(2.4)	(1.9)	(2.1)
	COVERT	M	18.1	18.8	19.4	18.8	18.4	19.4	18.8	18.6
		SD	(4.2)	(2.7)	(2.1)	(2.9)	(2.7)	(3.1)	(2.2)	(3.7)
COMPRE-HENSIVE TEST	OVERT	M	18.4	17.0	19.7	19.3	18.0	17.9	18.4	18.2
		SD	(4.9)	(5.9)	(4.2)	(4.5)	(5.0)	(4.4)	(3.6)	(4.2)
	COVERT	M	17.0	18.4	18.7	18.1	17.8	19.6	18.6	16.8
		SD	(6.2)	(5.1)	(5.2)	(5.0)	(4.4)	(5.5)	(3.7)	(5.4)
DELAYED TEST	OVERT	M	10.8	9.4	10.7	10.2	10.0	9.9	9.0	9.7
		SD	(4.8)	(6.3)	(4.8)	(4.8)	(5.7)	(4.6)	(4.4)	(4.7)
	COVERT	M	8.0	9.0	9.5	9.0	9.9	9.8	9.6	9.6
		SD	(5.9)	(4.4)	(5.5)	(5.4)	(5.4)	(5.3)	(3.9)	(4.8)

TABLE III-5

Combined Daily Tests: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure.

RESPONSE ACCURACY		SINGLE RESPONSE FRAMES				MULTIPLE RESPONSE FRAMES			
		OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Accurately Reproduced	M	15.9	14.8	12.1	12.5	17.4	15.3	14.3	12.9
	SD	(4.9)	(5.5)	(5.6)	(5.2)	(3.4)	(4.6)	(4.3)	(4.7)
0 to 1-Letter Spelling Inaccuracy	M	17.4	16.7	15.1	15.7	19.2	18.0	17.0	15.5
	SD	(4.0)	(5.1)	(5.9)	(5.0)	(2.7)	(4.0)	(3.7)	(4.8)
0 to 2-Letter Spelling Inaccuracy	M	17.9	17.0	15.5	16.5	19.4	18.3	17.3	16.1
	SD	(3.8)	(4.9)	(5.8)	(4.8)	(2.6)	(3.9)	(3.7)	(4.9)
0 to 5-Letter Spelling Inaccuracy	M	18.5	17.5	16.6	17.2	20.0	18.8	18.2	16.4
	SD	(3.4)	(4.8)	(5.3)	(4.5)	(2.2)	(3.4)	(2.9)	(4.8)

TABLE III-6

Combined Daily Tests: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Expanded Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure.

RESPONSE ACCURACY		SINGLE RESPONSE FRAMES				MULTIPLE RESPONSE FRAMES			
		OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Accurately Reproduced	M	16.7	14.4	13.4	13.6	16.9	18.0	13.4	13.4
	SD	(4.6)	(4.3)	(5.1)	(4.9)	(3.8)	(3.1)	(4.5)	(5.2)
0 to 1-Letter Spelling Inaccuracy	M	17.8	16.5	15.6	16.5	18.5	19.8	16.5	16.3
	SD	(4.0)	(4.2)	(4.1)	(4.8)	(3.2)	(2.5)	(3.7)	(5.1)
0 to 2-Letter Spelling Inaccuracy	M	18.1	16.9	16.2	16.9	19.0	19.9	17.1	17.0
	SD	(3.7)	(4.0)	(3.6)	(4.5)	(2.8)	(2.3)	(3.5)	(4.9)
0 to 5-Letter Spelling Inaccuracy	M	18.5	17.2	17.1	17.8	19.4	20.0	18.1	17.5
	SD	(3.4)	(3.5)	(3.1)	(4.1)	(2.4)	(2.3)	(3.0)	(4.7)

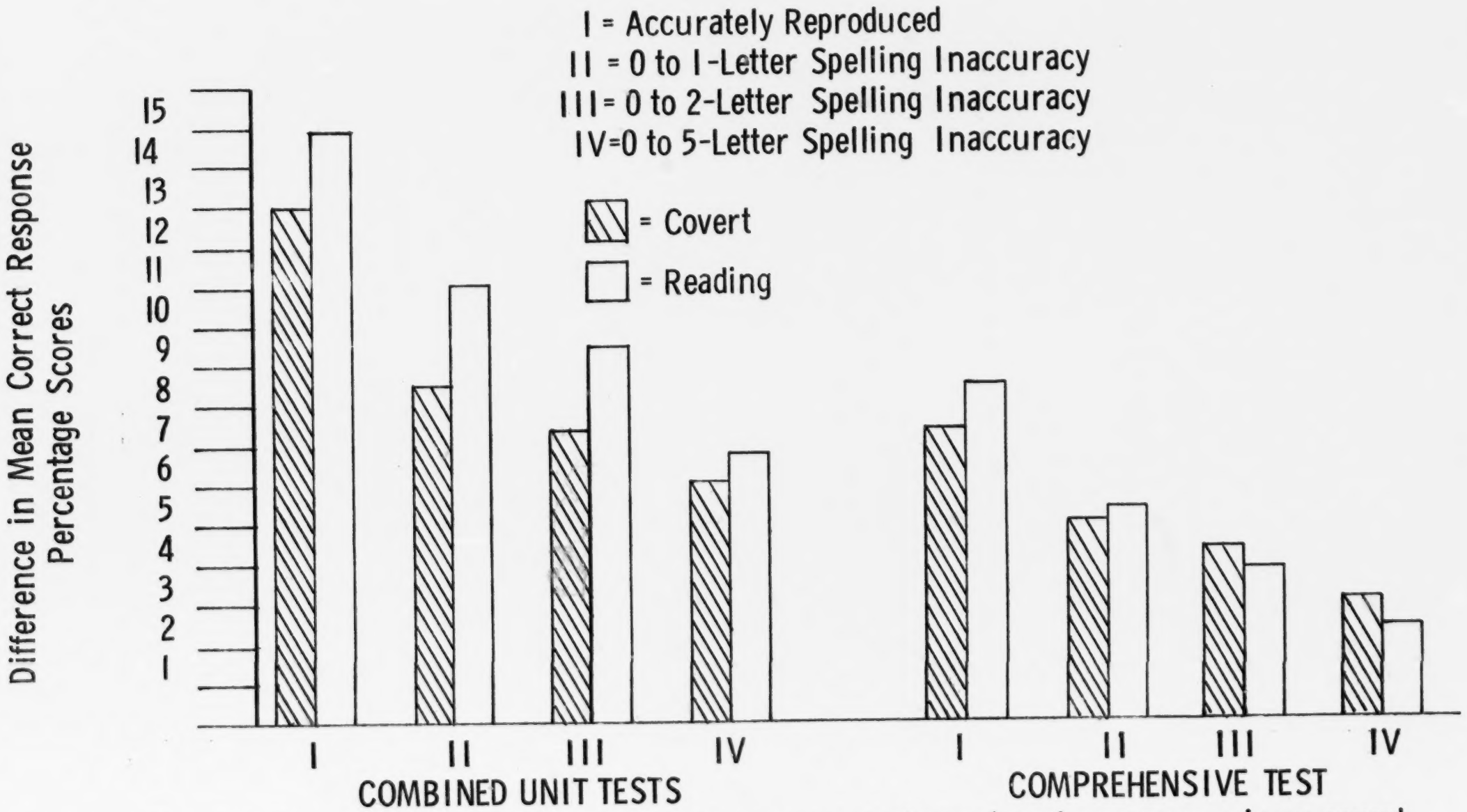

Fig. 8. Differences in mean correct response percentages (overt response minus covert response test scores, and overt response minus reading test scores) for two retention tests as a function of response reproduction accuracy.

TABLE III-7

Comprehensive Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure.

RESPONSE ACCURACY		SINGLE RESPONSE FRAMES				MULTIPLE RESPONSE FRAMES			
		OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Accurately Reproduced	M	13.9	12.9	12.0	12.0	16.2	14.9	13.6	12.1
	SD	(5.1)	(6.1)	(5.7)	(5.6)	(4.8)	(5.4)	(6.2)	(6.3)
0 to 1-Letter Spelling Inaccuracy	M	16.0	15.0	15.0	14.6	18.1	17.8	15.9	15.2
	SD	(5.2)	(6.0)	(5.7)	(5.8)	(4.6)	(5.1)	(6.2)	(6.2)
0 to 2-Letter Spelling Inaccuracy	M	16.2	15.7	15.7	15.3	18.4	18.2	16.6	15.8
	SD	(5.3)	(6.3)	(5.7)	(5.5)	(4.4)	(5.2)	(5.9)	(6.1)
0 to 5-Letter Spelling Inaccuracy	M	17.2	16.2	16.7	16.4	19.0	19.0	17.6	17.0
	SD	(5.1)	(6.2)	(5.7)	(5.5)	(4.1)	(4.8)	(5.6)	(6.1)

TABLE III-8

Comprehensive Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Expanded Frame Programs as a Function of Number of Responses Per Frame, Response Mode and Confirmation Procedure.

RESPONSE ACCURACY		SINGLE RESPONSE FRAMES				MULTIPLE RESPONSE FRAMES			
		OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Accurately Reproduced	M	14.1	14.4	12.7	13.5	15.5	15.0	12.4	11.9
	SD	(6.1)	(5.3)	(5.7)	(6.2)	(3.5)	(4.6)	(5.4)	(5.6)
0 to 1-Letter Spelling Inaccuracy	M	16.6	16.2	15.7	16.4	18.0	17.7	16.4	15.2
	SD	(5.6)	(5.3)	(5.2)	(6.2)	(3.8)	(3.9)	(4.7)	(5.3)
0 to 2-Letter Spelling Inaccuracy	M	17.1	16.7	16.5	17.0	18.4	18.4	17.0	15.9
	SD	(5.4)	(5.1)	(5.2)	(5.9)	(3.9)	(4.0)	(4.3)	(5.3)
0 to 5-Letter Spelling Inaccuracy	M	17.9	17.3	17.2	17.6	18.8	18.9	18.1	17.1
	SD	(5.2)	(5.2)	(5.0)	(5.6)	(3.8)	(3.9)	(4.2)	(5.0)

TABLE III-9

Delayed Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic Frame Programs as a Function of Number of Responses Required Per Frame, Response Mode and Confirmation Procedure.

RESPONSE ACCURACY		SINGLE RESPONSE FRAMES				MULTIPLE RESPONSE FRAMES			
		OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Accurately Reproduced	M	5.0	4.8	3.6	4.1	4.8	4.3	4.6	4.5
	SD	(3.6)	(4.3)	(4.8)	(3.8)	(4.2)	(4.7)	(4.7)	(4.5)
0 to 1-Letter Spelling Inaccuracy	M	5.9	6.3	4.7	5.3	6.2	5.8	6.5	5.7
	SD	(3.6)	(5.8)	(5.4)	(4.4)	(4.6)	(5.6)	(6.0)	(5.1)
0 to 2-Letter Spelling Inaccuracy	M	6.3	6.7	5.2	5.9	6.5	6.6	7.2	6.3
	SD	(3.7)	(6.1)	(5.8)	(4.5)	(4.7)	(5.8)	(6.2)	(5.2)
0 to 5-Letter Spelling Inaccuracy	M	6.7	7.3	5.8	6.8	7.2	7.6	7.8	7.0
	SD	(3.8)	(5.9)	(5.9)	(4.5)	(5.2)	(5.9)	(6.6)	(5.2)

TABLE III-10

Delayed Test: Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Expanded Frame Programs as a Function of Number of Responses Required Per Frame, Response Mode and Confirmation Procedure.

RESPONSE ACCURACY		SINGLE RESPONSE FRAMES				MULTIPLE RESPONSE FRAMES			
		OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Accurately Reproduced	M	4.9	4.5	4.3	4.8	3.5	3.6	2.6	3.8
	SD	(3.9)	(3.5)	(4.8)	(4.5)	(3.5)	(2.6)	(2.5)	(2.8)
0 to 1-Letter Spelling Inaccuracy	M	5.7	5.8	5.8	6.6	4.7	5.0	5.1	5.4
	SD	(4.3)	(4.1)	(5.6)	(5.4)	(4.0)	(3.5)	(4.3)	(4.3)
0 to 2-Letter Spelling Inaccuracy	M	6.0	6.3	6.3	6.9	5.3	5.8	4.7	5.7
	SD	(4.4)	(4.1)	(5.8)	(5.3)	(4.0)	(3.6)	(3.1)	(3.9)
0 to 5-Letter Spelling Inaccuracy	M	6.8	6.8	6.9	7.7	6.0	6.5	5.6	6.7
	SD	(4.7)	(4.1)	(5.8)	(5.7)	(4.1)	(4.1)	(3.5)	(4.0)

TABLE III-11

Summary of Analyses of Variance of Correctly Recalled and Accurately Reproduced Definition Test Item Responses for Three Retention Tests

Source	df	Combined Unit Tests		Comprehensive Test		Delayed Test	
		MS	F	MS	F	MS	F
Frame Size (FS)	1	3.61	----	2.40	----	7.02	----
Number of Responses (NR)	1	7.29	----	12.60	----	0.06	----
Response Mode (RM)	1	62.41	5.85*	85.56	3.48	49.70	2.14
Confirmation (C)	1	1.21	----	3.42	----	7.56	----
FS x NR	1	6.25	----	76.56	3.11	37.82	1.63
FS x RM	1	0.01	----	8.12	----	30.80	1.33
FS x C	1	2.89	----	0.42	----	9.30	----
NR x RM	1	16.81	1.57	49.72	2.02	3.06	----
NR x C	1	0.81	----	25.50	1.04	9.30	----
RM x C	1	20.25	1.90	11.22	----	17.22	----
FS x NR x RM	1	18.49	1.73	0.01	----	0.04	----
FS x NR x C	1	0.81	----	3.80	----	0.06	----
NR x RM x C	1	9.61	----	69.06	----	9.30	----
FS x RM x C	1	4.41	----	2.40	----	10.56	----
FS x NR x RM x C	1	2.25	----	0.42	----	14.06	----
Within	384	10.67		24.62		23.22	

* $p < .05$

TABLE III-12

Summary of Analyses of Variance of Correctly Recalled Definition Test Item Responses Scored Without Regard for Spelling Accuracy for Three Retention Tests

Source	df	Combined Unit Tests		Comprehensive Test		Delayed Test	
		MS	F	MS	F	MS	F
Frame Size (FS)	1	0.64	----	2.72	----	1.10	----
Number of Responses (NR)	1	30.25	3.58	22.56	----	0.30	----
Response Mode (RM)	1	2.89	----	6.50	----	45.56	1.76
Confirmation (C)	1	2.25	----	2.10	----	1.82	----
FS x NR	1	6.25	----	66.42	2.80	23.52	----
FS x RM	1	0.81	----	9.92	----	49.70	1.92
FS x C	1	5.29	----	1.10	----	6.00	----
NR x RM	1	9.00	1.06	39.06	1.65	3.06	----
NR x C	1	1.00	----	35.40	1.49	0.30	----
RM x C	1	11.56	1.37	11.22	----	5.06	----
FS x RM x RM	1	3.24	----	0.56	----	0.01	----
FS x NR x C	1	0.16	----	10.56	----	3.80	----
NR x RM x C	1	28.09	3.32	66.42	2.80	14.82	----
FS x RM x C	1	0.64	----	8.70	----	15.60	----
FS x NR x RM x C	1	0.25	----	0.42	----	5.52	----
Within	384	8.46		23.73		25.82	

TABLE III-13

Summary of Analyses of Variance of Medical Term Test Item Responses by Response Accuracy Classification for the Combined Unit Tests.

Source	df	Accurately Reproduced		0 to 1-Letter Spelling Inaccuracy		0 to 2-Letter Spelling Inaccuracy		0 to 5-Letter Spelling Inaccuracy	
		MS	F	MS	F	MS	F	MS	F
Frame Size (FS)	1	31.92	1.46	12.96	-----	14.44	-----	8.41	-----
Number of Responses (NR)	1	109.20	5.00*	134.56	7.41**	125.44	7.52**	96.04	6.94**
Response Mode (RM)	1	897.00	41.07***	384.16	21.14***	289.00	17.34***	193.21	13.96***
Confirmation (C)	1	42.90	1.96	7.84	-----	5.76	-----	22.09	1.60
FS x NR	1	1.32	-----	0.01	-----	0.64	-----	0.81	-----
FS x RM	1	0.56	-----	0.09	-----	0.36	-----	4.84	-----
FS x C	1	16.40	-----	18.49	1.02	9.00	-----	12.96	-----
NR x RM	1	18.92	-----	34.81	1.92	29.16	1.75	39.69	2.87
NR x C	1	0.42	-----	1.69	-----	1.96	-----	6.25	-----
RM x C	1	19.80	-----	5.29	-----	10.24	-----	4.84	-----
RS x NR x RM	1	33.06	1.51	5.76	-----	4.00	-----	1.44	-----
FS x NR x C	1	55.50	2.54	25.00	1.38	21.16	1.27	16.00	1.16
NR x RM x C	1	28.62	1.31	40.96	2.25	38.44	2.31	49.00	3.54
FS x RM x C	1	1.32	-----	0.36	-----	1.44	-----	0.09	-----
FS x NR x RM x C	1	12.60	-----	6.25	-----	1.44	-----	1.69	-----
Within	384	21.84		18.17		16.67		13.84	

* p <.05
** p <.01
*** p <.001

TABLE III-14

Summary of Analyses of Variance of Medical Term Test Item Responses by Response Accuracy Classification for the Comprehensive Tests.

Source	df	Accurately Reproduced		0 to 1-Letter Spelling Inaccuracy		0 to 2-Letter Spelling Inaccuracy		0 to 5-Letter Spelling Inaccuracy	
		MS	F	MS	F	MS	F	MS	F
Frame Size (FS)	1	5.06	-----	33.06	1.15	37.82	1.37	24.01	-----
Number of Responses (NR)	1	61.62	2.02	122.10	4.26*	111.30	4.03*	127.69	4.90*
Response Mode (RM)	1	434.72	14.31***	183.60	6.41*	131.10	4.74*	68.89	2.64
Confirmation (C)	1	22.56	-----	17.22	-----	11.90	-----	14.43	-----
FS x NR	1	55.50	1.82	28.62	1.00	23.52	-----	15.21	-----
FS x RM	1	0.12	-----	3.42	-----	1.32	-----	1.21	-----
FS x C	1	24.50	-----	2.72	-----	1.82	-----	1.00	-----
NR x RM	1	63.20	2.08	71.40	2.49	71.40	2.58	44.89	1.72
NR x C	1	25.50	-----	2.72	-----	2.72	-----	0.00	-----
RM x C	1	2.72	-----	0.30	-----	0.72	-----	0.04	-----
FS x NR x RM	1	2.10	-----	0.00	-----	0.20	-----	1.69	-----
FS x NR x C	1	0.03	-----	8.12	-----	2.10	-----	3.24	-----
NR x RM x C	1	4.20	-----	14.06	-----	10.56	-----	17.64	-----
FS x RM x C	1	0.20	-----	0.00	-----	0.30	-----	0.03	-----
FS x NR x RM x C	1	1.10	-----	0.90	-----	2.72	-----	1.44	-----
Within	384	30.38		28.63		27.62		26.06	

* p <.05
*** p <.001

Table III-14, containing the results of the comprehensive test data analyses, reflects the effects of the converging means and demonstrates that the comparison of the effects of response mode is highly dependent upon the criterion used to score the test item responses. The effects of Response Mode are clearly significant when responses are scored with the stringent "accurately reproduced" criterion. However, when overt and covert responding are compared in terms of the most lenient criterion (0 to 5-letter spelling inaccuracy), the superiority of the overt response is no longer significant. A very small part of the convergence effect on the combined unit tests can be attributed to the ceiling created by the limited number of items on these tests. While none of the subjects in the covert response and reading groups attained a maximal score when the "accurately reproduced" criterion was used to score the daily tests, six of the overt response subjects were able to achieve this score and were consequently unable to show any further improvement when lesser degrees of spelling accuracy were considered. On the comprehensive test, however, none of the subjects in any of the response mode groups obtained a maximal score with any scoring criterion.

The results entered in Table III-14 also demonstrate significant main effects for Number of Responses. As with the combined unit tests, the Multiple Response Frame Programs yielded higher comprehensive test scores for medical term responses. Significant differences, however, were limited to the three classifications for inaccurately reproduced response. No other significant main or interaction effects were revealed by the analyses of the comprehensive test medical term responses.

Table III-15 contains a summary of the results obtained from the analyses performed on the delayed retention test. As can be seen, Response Mode and Number of Responses were no longer significant sources of variation, and all other effects were nonsignificant as well.

The mean retention test score for each of the two groups instructed by the reading programs is presented in Table III-16 (definition test item responses) and Table III-17 (medical term test item responses). The differences observed between the scores of the Basic and Expanded Frame versions were examined by t tests. All comparisons yielded nonsignificant results (p's >.05); that is, no statistically reliable differences were found between the scores of the two frame size groups on any of the retention tests, whether for definition or medical term responses at any of the accuracy of response classifications.

Dunnett's test (Winer, 1962) was used to compare the test performance results of the reading program subjects with the results obtained from the subjects in the overt and covert response groups. In this analysis the Basic and Expanded Frame Program test scores were combined and contrasted with the scores from each of the response mode groups collapsed across the frame size, number of responses per frame and confirmation variations.

No significant differences at the .05 level between reading and overt or reading and covert responding were found for definition test item responses on any of the retention tests, whether for "accurately reproduced" responses or responses scored without consideration for

TABLE III-15

Summary of Analyses of Variance of Medical Term Test Item Responses by Response Accuracy Classification for the Delayed Tests.

Source	df	Accurately Reproduced		0 to 1-Letter Spelling Inaccuracy		0 to 2-Letter Spelling Inaccuracy		0 to 5-Letter Spelling Inaccuracy	
		MS	F	MS	F	MS	F	MS	F
Frame Size (FS)	1	22.09	1.38	23.52	1.05	19.80	-----	14.06	-----
Number of Responses (NR)	1	27.04	1.69	14.82	-----	3.42	-----	0.12	-----
Response Mode (RM)	1	14.44	-----	3.80	-----	2.10	-----	0.32	-----
Confirmation (C)	1	1.44	-----	6.50	-----	11.90	-----	19.80	-----
FS x NR	1	51.84	3.25	78.32	3.50	64.80	2.74	58.52	2.31
FS x RM	1	1.44	-----	8.12	-----	5.52	-----	8.70	-----
FS x C	1	4.00	-----	7.56	-----	6.00	-----	2.40	-----
NR x RM	1	4.41	-----	1.56	-----	0.90	-----	0.12	-----
NR x C	1	0.09	-----	5.06	-----	3.06	-----	2.10	-----
RM x C	1	15.21	-----	2.10	-----	0.20	-----	0.56	-----
FS x NR x RM	1	7.29	-----	24.50	1.90	23.52	-----	11.22	-----
FS x NR x C	1	6.25	-----	11.22	-----	10.56	-----	10.56	-----
NR x RM x C	1	0.00	-----	0.56	-----	1.32	-----	3.80	-----
FS x RM x C	1	1.21	-----	4.20	-----	3.42	-----	7.02	-----
FS x NR x RM x C	1	0.36	-----	0.72	-----	3.42	-----	2.40	-----
Within	384	15.93		22.36		23.60		25.29	

TABLE III-16

Means and Standard Deviations of Correctly Recalled Definition Test Item Responses by the Basic and the Expanded Frame Reading Program Subjects Scored With (Criterion I) and Without Regard (Criterion II) for Spelling Accuracy for Three Retention Tests.

Retention Test	Frame Size	Criterion I		Criterion II	
		M	SD	M	SD
Combined Unit Tests	Basic	17.6	(2.9)	19.6	(2.5)
	Expanded	17.0	(3.0)	18.9	(2.6)
Comprehensive Test	Basic	17.2	(5.4)	18.1	(5.4)
	Expanded	17.8	(4.8)	18.8	(5.0)
Delayed Test	Basic	9.6	(5.0)	10.2	(5.3)
	Expanded	9.2	(5.0)	10.0	(4.7)

TABLE III-17

Mean Number of Medical Term Responses, by Response Accuracy Classification, for the Basic and Expanded Frame Reading Programs on Three Retention Tests.

RETENTION TEST	FRAME SIZE	Accurately Reproduced M	Accurately Reproduced SD	0 to 1-Letter Spelling Inaccuracy M	0 to 1-Letter Spelling Inaccuracy SD	0 to 2-Letter Spelling Inaccuracy M	0 to 2-Letter Spelling Inaccuracy SD	0 to 5-Letter Spelling Inaccuracy M	0 to 5-Letter Spelling Inaccuracy SD
COMBINED UNIT TESTS	BASIC	12.7	(5.7)	15.4	(4.9)	16.2	(4.7)	17.4	(4.0)
	EXPANDED	12.7	(5.1)	15.5	(4.3)	16.0	(4.1)	17.0	(3.6)
COMPRE-HENSIVE TEST	BASIC	12.6	(5.8)	15.8	(5.6)	16.6	(5.6)	17.6	(5.6)
	EXPANDED	11.8	(5.5)	15.0	(5.1)	16.0	(5.1)	17.2	(4.9)
DELAYED TEST	BASIC	3.7	(3.5)	5.1	(4.3)	5.8	(4.5)	6.7	(4.6)
	EXPANDED	4.6	(4.1)	5.4	(4.3)	6.0	(4.3)	6.8	(4.4)

Table III-18

Means and Standard Deviations of the Types of Program Errors Made in the Frame Responses Elicited in Common by the Single and Multiple Response Frame Programs as a Function of Frame Size and Confirmation Procedure.

PROGRAM ERROR CLASSIFICATION		SINGLE RESPONSE FRAME PROGRAMS				MULTIPLE RESPONSE FRAME PROGRAMS			
		BASIC FRAMES		EXPANDED FRAMES		BASIC FRAMES		EXPANDED FRAMES	
		CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
Medical Term Errors	M	7.8	10.6	9.6	12.4	7.5	10.6	7.8	7.2
	SD	(6.6)	(14.6)	(7.1)	(13.3)	(5.7)	(15.6)	(5.0)	(6.0)
Correct Medical Terms with Spelling Errors	M	11.8	15.3	13.4	15.4	15.9	18.9	14.6	14.3
	SD	(8.5)	(13.7)	(9.3)	(7.7)	(11.5)	(15.3)	(10.7)	(11.0)
Definition Errors	M	4.8	6.0	4.7	5.1	5.1	4.4	4.5	4.3
	SD	(4.3)	(6.5)	(3.1)	(4.3)	(4.1)	(4.2)	(3.2)	(3.1)
Correct Definitions with Spelling Errors	M	3.6	4.6	3.7	2.8	5.4	5.0	4.0	4.9
	SD	(3.7)	(2.6)	(2.6)	(2.3)	(5.4)	(4.1)	(3.4)	(3.9)

spelling errors. For medical term test item responses, the reading program test scores closely paralleled the covert response scores in two respects. First, Dunnett's test did not reveal any significant differences at the .05 level between reading and covert responding at any response accuracy classification for any of the retention tests. Secondly, analogous to comparisons between covert and overt response test scores, the differences between the reading and the overt response test scores became smaller when less emphasis was placed upon accurate reproduction in scoring the responses. This is represented graphically in Fig. 8. For this comparison the mean number of responses scored by each accuracy of response classification was converted into a mean correct response percentage score. The comparative differences were derived by subtracting the reading and covert response mean scores from the overt response mean score. Data from the delayed retention test are omitted since Dunnett's test indicated that the differences in the test scores between reading and overt responding, as well as between reading and covert responding, were not significant (p's >.05). On the combined unit tests, the difference between reading and overt responding was statistically significant at the .005 level for all response accuracy classifications. On the comprehensive test, overt responding produced significantly higher scores only for "accurately reproduced" responses ($p < .005$), and for "0 to 1-letter spelling inaccuracy" responses ($p < .05$).

Program Performance

The program responses made by the subjects in the overt response groups were analyzed to determine the effects of the variations in frame size, number of responses per frame and confirmation on program performance. Error rates for each subject were calculated independently for each of the specific kinds of responses elicited by the program frames, that is, for medical terms, definitions and instructional terms. The type of error made for each response classification was also considered. A response was designated as an error when subjects either: (a) provided a patently wrong response, (b) did not fully complete the required response, or (c) were unable to respond at all. A separate tabulation was made of responses that were misspelled but were nevertheless judged to be recognizable as the required responses. The three judges who evaluated the retention test data were also used to score the program responses. Medical term and definition responses were scored by the same criteria in both evaluations.

The left side of Table III-18 provides a comparison of frame size and confirmation effects on Single Response Frame Program performance. Data for the 38 instructional term responses are not separately tabulated in this table. Since very few incorrect responses or misspellings were in evidence with these terms, they were combined with the definition responses for the analysis.

The inclusion of the Multiple Response Frame Program data on the right side of Table III-18 requires further explanation. As previously described, the Single Response Frame versions of the programs were developed by filling in beforehand all but one of the response blanks in 346 frames of the Multiple Response Frame versions. In addition,

each of the remaining 38 frames in the Multiple Response Frame versions was divided into two parts to create more criterion frames for the Single Response Frame Programs. Consequently, the comparison between the Single and Multiple Response Frame Programs in the table involves the 244 responses which were elicited by the same frame material in both versions of the programs.

Summaries of the 2 x 2 x 2 analyses of variance conducted to determine the effects of the program variations on each type of program error appear in Table III-19. The "correct medical terms with spelling errors" classification includes responses that were scored as 1 to 5-letter spelling inaccuracies. As can be seen the verbiage added to the Basic Frame Program frames had no effect on program performance. The only significant effect created by increasing the number of required frame responses was an increase in the number of correct, but misspelled, definitions. Although an examination of the means for Single Response Frame Program reveals that in almost every comparison the absence of confirmation is associated with an increase in program errors, none of the analyses of the various types of errors resulted in either a significant main effect or a significant interaction effect involving confirmation.

The program errors made by the subjects instructed by the Multiple Response Frame Programs are categorized in Table III-20. A 2 x 2 analysis of variance was used to evaluate the effects of frame size and confirmation on each type of program error. Table III-21, which contains the results of these analyses, shows that neither variable was found to be a significant source of variation. It is observed that Confirmation and Frame Size X Confirmation produced F ratios that were all less than unity.

Product-moment correlation coefficients were computed to determine the relationship between program performance and retention test scores. A distinction was again made between program errors which indicated a subject's inability to provide the appropriate responses and errors which were the result of inaccurate reproduction. Program errors on instructional term responses were excluded from this analysis.

The correlation coefficients between retention test scores and number of program responses which were either incorrect, incomplete or omitted are presented in Table III-22. Table III-23 contains the results obtained when retention test scores were correlated with the number of misspelled medical term program responses (1 to 5-letter spelling inaccuracies), and the number of correct definition program responses which contained misspelled words. Only the coefficients derived when program errors were correlated with test scores based upon the number of "accurately reproduced" responses were entered into these tables. While not presented, the same results were obtained for test scores which included recognizable but inaccurately reproduced responses across all medical term and definition response accuracy classifications.

These findings indicate that program performance is, indeed, a valid predictor of post program achievement. Low program error rates are significantly associated with high retention test scores. Both

TABLE III-19

Summary of the Analyses of Variance of the Types of Program Errors in the Frame Responses Elicited in Common by the Single and Multiple Frame Programs.

SOURCE	df	MEDICAL TERM ERRORS		CORRECT MEDICAL TERMS WITH SPELLING ERRORS		DEFINITION ERRORS		CORRECT DEFINITIONS WITH SPELLING ERRORS	
		MS	F	MS	F	MS	F	MS	F
Frame Size (FS)	1	0.98	----	54.08	----	10.13	----	30.42	2.32
Number of Responses (NR)	1	169.28	1.65	184.32	1.47	15.13	----	67.28	5.13*
Confirmation (C)	1	208.08	2.03	212.18	1.69	1.81	----	0.98	----
FS x NR	1	141.12	1.38	184.32	1.47	0.41	----	0.08	----
FS x C	1	42.32	----	74.42	----	0.41	----	0.98	----
NR x C	1	30.42	----	4.60	----	19.85	1.12	0.08	----
FS x NR x C	1	36.98	----	8.82	----	5.45	----	32.00	2.44
Within	192	102.57		125.60		17.76		13.11	

TABLE III-20

Means and Standard Deviations of the Types of Program Errors Made by Subjects Instructed by the Multiple Response Frame Programs as a Function of Frame Size and Confirmation Procedure.

		BASIC FRAME PROGRAMS		EXPANDED FRAME PROGRAMS	
		Confirmation	No Confirmation	Confirmation	No Confirmation
Medical Term Errors	M	16.6	21.0	14.5	14.2
	SD	(12.0)	(24.4)	(9.6)	(11.3)
Correct Medical Terms with Spelling Errors	M	32.3	35.1	30.2	27.0
	SD	(23.0)	(29.9)	(19.6)	(20.9)
Definition Errors	M	14.5	15.0	13.4	14.9
	SD	(9.8)	(15.6)	(8.2)	(11.3)
Correct Definitions with Spelling Errors	M	20.4	22.0	16.7	20.9
	SD	(19.8)	(17.5)	(12.2)	(15.7)
Instructional Term Errors	M	2.0	2.9	2.2	2.2
	SD	(1.5)	(4.3)	(1.8)	(2.2)
Correct Instructional Terms with Spelling Errors	M	1.3	1.4	0.7	1.0
	SD	(1.9)	(2.0)	(1.4)	(1.8)

TABLE III-21

Summary of the Analyses of Variance of the Types of Program Errors Made by Subjects Instructed by the Multiple Response Frame Programs.

Source	df	Medical Term Errors		Correct Medical Terms with Spelling Errors		Definition Errors		Correct Definition with Spelling Errors		Instructional Term Errors		Correct Instructional Terms with Spelling Errors	
		MS	F	MS	F	MS	F	MS	F	MS	F	MS	F
Frame Size (FS)	1	506.25	2.11	650.25	1.16	7.84	----	146.41	----	2.25	----	5.29	1.67
Confirmation (C)	1	106.09	----	1.21	----	25.00	----	204.49	----	5.29	----	1.21	----
FS x C	1	146.41	----	222.01	----	6.76	----	39.69	----	4.41	----	0.21	----
Within	96	239.92		562.60		134.32		272.84		7.27		3.16	

TABLE III-22

Product-Moment Correlations* Between Program Errors (Incorrect and Incomplete Responses and Omissions) and Retention Test Performance.

RETENTION TEST SCORES		MEDICAL TERM PROGRAM ERRORS								DEFINITION PROGRAM ERRORS							
		BASIC FRAME				EXPANDED FRAME				BASIC FRAME				EXPANDED FRAME			
		SINGLE RESPONSE		MULTIPLE RESPONSE		SINGLE RESPONSE		MULTIPLE RESPONSE		SINGLE RESPONSE		MULTIPLE RESPONSE		SINGLE RESPONSE		MULTIPLE RESPONSE	
		C	NC	C	NC	C	NC	C	NC	C	NC	C	NC	C	NC	C	NC
COMBINED DAILY TESTS	MEDICAL TERM ITEMS	.77	.66	.65	.82	.78	.77	.72	.55	.44	.68	.56	.67	.35	.75	.58	.59
	DEFINITION ITEMS	.82	.79	.64	.72	.58	.80	.75	.38	.59	.80	.61	.80	.47	.75	.60	.40
COMPREHENSIVE TEST	MEDICAL TERM ITEMS	.69	.63	.65	.64	.71	.56	.39	.43	.39	.71	.54	.65	.61	.50	.56	.59
	DEFINITION ITEMS	.79	.56	.57	.72	.72	.65	.33	.16	.47	.69	.42	.77	.64	.63	.49	.46
DELAYED TEST	MEDICAL TERM ITEMS	.61	.32	.40	.32	.45	.33	.27	.45	.43	.40	.31	.30	.49	.40	.23	.56
	DEFINITION ITEMS	.70	.40	.41	.38	.44	.60	.48	.32	.50	.54	.38	.47	.38	.57	.49	.49

* Each value represents a negative correlation coefficient.

Confidence levels for N = 25 (one-tailed test): $p \leq .05 = .33$, $p \leq .01 = .45$.

TABLE III-23

Product-Moment Correlations* Between Program Errors (1 to 5-Letter Spelling Inaccuracy Medical Term Responses and Correct Definitions Containing Misspelled Words) and Retention Test Performance.

RETENTION TEST SCORES		MEDICAL TERM PROGRAM ERRORS								DEFINITION PROGRAM ERRORS							
		BASIC FRAME				EXPANDED FRAME				BASIC FRAME				EXPANDED FRAME			
		SINGLE RESPONSE		MULTIPLE RESPONSE		SINGLE RESPONSE		MULTIPLE RESPONSE		SINGLE RESPONSE		MULTIPLE RESPONSE		SINGLE RESPONSE		MULTIPLE RESPONSE	
		C	NC	C	NC	C	NC	C	NC	C	NC	C	NC	C	NC	C	NC
COMBINED DAILY TESTS	MEDICAL TERM ITEMS	.76	.60	.65	.77	.66	.76	.58	.58	.52	.46	.53	.58	.47	.23	.47	.23
	DEFINITION ITEMS	.81	.73	.72	.68	.49	.71	.39	.43	.65	.38	.66	.60	.31	.29	.57	.62
COMPREHENSIVE TEST	MEDICAL TERM ITEMS	.70	.66	.64	.68	.65	.65	.60	.71	.59	.47	.50	.61	.50	.27	.44	.28
	DEFINITION ITEMS	.77	.54	.68	.70	.62	.63	.24	.48	.66	.35	.55	.61	.47	.20	.34	.44
DELAYED TEST	MEDICAL TERM ITEMS	.59	.37	.48	.35	.42	.53	.18	.49	.57	.17	.27	.26	.39	.44	.23	.16
	DEFINITION ITEMS	.68	.40	.49	.28	.38	.51	.29	.32	.57	.27	.38	.44	.30	.42	.40	.18

* Each value respresents a negative correlation coefficient.

Confidence levels for N = 25 (one-tailed test): $p \leq .05 = .33$, $p \leq .01 = .45$.

TABLE III-24

Means and Standard Deviations of Test Completion Times in Minutes for Three Retention Tests as a Function of the Program Variations in Frame Size, Number of Responses Per Frame, Response Mode and Confirmation Procedure.

RETENTION TEST	FRAME SIZE		SINGLE RESPONSE FRAME PROGRAMS				MULTIPLE RESPONSE FRAME PROGRAMS			
			OVERT RESPONSE		COVERT RESPONSE		OVERT RESPONSE		COVERT RESPONSE	
			CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.	CONF.	NO CONF.
COMBINED DAILY TESTS	BASIC	M	22.2	23.0	25.4	26.2	21.1	21.6	25.2	27.6
		SD	(4.9)	(6.7)	(8.3)	(8.9)	(5.6)	(7.3)	(7.3)	(8.1)
	EXPANDED	M	20.6	21.4	25.7	24.8	18.9	19.8	25.4	27.7
		SD	(6.0)	(4.7)	(6.3)	(8.2)	(4.9)	(4.6)	(7.8)	(8.9)
COMPREHENSIVE TEST	BASIC	M	23.7	22.8	23.6	24.4	22.6	23.1	26.8	25.2
		SD	(7.2)	(8.2)	(7.3)	(7.8)	(8.6)	(6.4)	(10.3)	(6.5)
	EXPANDED	M	21.0	20.8	25.8	24.5	21.3	22.4	24.3	26.2
		SD	(7.1)	(4.4)	(7.7)	(6.9)	(5.0)	(5.4)	(6.1)	(7.5)
DELAYED TEST	BASIC	M	24.3	20.8	19.5	24.5	20.9	20.5	22.8	23.4
		SD	(4.7)	(8.5)	(6.3)	(7.4)	(5.9)	(6.3)	(10.9)	(7.0)
	EXPANDED	M	21.8	18.7	22.2	22.2	20.3	20.3	21.4	23.5
		SD	(7.2)	(5.2)	(5.2)	(5.2)	(5.9)	(6.9)	(6.7)	(6.7)

TABLE III-25

Summary of Analyses of Variance of Test Completion Times for Three Retention Tests.

Source	df	Combined Unit Tests		Comprehensive Test		Delayed Test	
		MS	F	MS	F	MS	F
Frame Size (FS)	1	101.00	2.10	55.50	1.08	60.84	1.32
Number of Responses (NR)	1	7.02	-----	45.56	-----	1.21	-----
Response Mode (RM)	1	2,425.56	50.43***	820.82	15.97***	219.04	4.76*
Confirmation (C)	1	93.12	1.94	0.12	-----	0.64	-----
FS x NR	1	0.72	-----	1.32	-----	7.29	-----
FS x RM	1	63.20	1.31	87.42	1.70	33.64	-----
FS x C	1	3.42	-----	11.90	-----	12.96	-----
NR x RM	1	147.62	3.07	14.06	-----	59.29	1.29
NR x C	1	33.06	-----	18.06	-----	24.01	-----
RM x C	1	4.20	-----	0.42	-----	338.56	7.36**
FS x NR x RM	1	9.30	-----	66.42	1.29	44.89	-----
FS x NR x C	1	6.00	-----	51.11	-----	68.89	1.50
NR x RM x C	1	40.32	-----	6.50	-----	110.25	2.40
FS x RM x C	1	3.12	-----	0.01	-----	29.16	-----
FS x NR x RM x C	1	2.72	-----	51.12	-----	68.89	1.50
Within	384	48.10		51.40		46.00	

* p <.05

** p <.01

*** p <.001

TABLE III-26

Means and Standard Deviations of Test Completion Times in Minutes for the Subjects Instructed by the Basic and Expanded Frame Reading Programs on Three Retention Tests.

Retention Test	Basic Frame Program		Expanded Frame Program	
	M	SD	M	SD
Combined Unit Tests	26.0	(8.0)	28.7	(9.4)
Comprehensive Test	24.6	(6.3)	25.7	(6.3)
Delayed Test	24.0	(7.7)	23.7	(7.0)

medical term and definition errors can be seen to be equally reliable as predictors of retention test performance regardless of whether the test scores are based on medical term or definition responses. In addition, the criteria used to identify program errors, whether simply correctness of recall or accuracy of reproduction, would appear to have no differential effect on the correlations.

The correlations are generally highest for the combined daily tests and the comprehensive test results. Beyond the pattern evident in the decline in correlation coefficient values from immediate to long term retention, no other consistent relationship is apparent. None of the program variations - basic or expanded frame, single or multiple response, and confirmation and nonconfirmation - appears to have any greater or lesser influence on the predictability of post-program achievement based on program errors.

Test Completion Times

Table III-24 presents the means and standard deviations of the test completion times in minutes for each of the program variations, except for the two reading versions. The combined unit test entries represent the mean total time required to complete all of the four daily tests but do not include the time taken for the essay sections of the first three daily tests.

Examination of the tabulation of the time data for the combined unit tests reveals that all of the covert response groups had longer mean completion times than any of the overt response groups. With one exception, this is also true for the comprehensive test data. Although there is some overlap among the response mode groups on the delayed retention test, the covert response mode groups generally display the longer mean completion times.

Results of four-way analyses of variance of the time data (Table III-25) showed that the differences attributable to response mode were significant for all three retention tests. The only other significant effect observed was the Response Mode x Confirmation interaction on the delayed retention test. An analysis of the groups involved in this interaction by Dunn's test revealed that the differences between the overt and covert response groups which received confirmation were not significant. Under conditions of nonconfirmation, however, the covert response subjects showed a significantly longer mean test completion time ($p < .05$). On the combined unit tests and the comprehensive test this interactive effect produced F ratios that were less than unity.

The test completion times of the groups instructed by the reading programs are shown in Table III-26. The differences between the Basic and Expanded Frame versions observed in this table were examined by t tests and found to be nonsignificant for all three retention tests. The two frame size groups were subsequently combined and comparisons were made between the mean test completion times of all the reading program subjects and the subjects instructed by the overt and the covert response programs. Dunnett's test was used for these comparisons.

No significant differences at the .05 level were found between the mean times of all the subjects who were provided with reading programs and subjects with programs which required covert responses. Subjects instructed by overt response programs showed shorter mean completion times than reading program subjects on both the combined unit tests ($p < .005$) and the comprehensive test ($p < .01$). Comparisons on the delayed test were made between the completion times of the reading program subjects and the subjects in the groups involved in the significant Response Mode x Confirmation interaction. Reading program subjects took significantly longer to complete the test than either the overt response/no confirmation subjects ($p < .005$) and the covert response/confirmation subjects ($p < .05$). The other comparisons yielded non-significant differences.

Discussion

When programmed instruction emerged from the operant conditioning laboratory, it was presented to the potential user complete with a behavioral paradigm and a definitive set of features. All of these features represented the functional counterparts of animal conditioning techniques and as such were considered indispensible. Behavioral considerations dictated the necessity of following certain specifications in developing self-instructional materials. One of these specifications placed restrictions on the informational content of program frames. It was proposed that limiting frames specifically to the critical content necessary for eliciting the desired responses would aid both program learning as well as subsequent recall. Ostensibly, frames constructed in this manner would enable the learner to make the appropriate program responses with a low probability of error, and also insure that the associations formed between the learner's responses and the contextual frame material would not be interferred with by irrelevant frame content.

The present findings did not provide any empirical support for the alleged merits of restricted frame content. Adding irrelevant material to the content of each frame in the Basic Frame Programs to create the Expanded Frame versions had no significant effect on either program errors or retention test performance. The only reliable effect attributable to frame size was the significantly longer time taken by subjects to complete the Expanded Frame Programs. Contrary to the expectations of Kemp and Holland (1966), no significant interaction involving response mode and amount of frame content was observed.

The evidence gathered on the effects of confirmation is also difficult to reconcile with an additional specification proposed for the development of programmed materials: the provision for a reinforcement procedure. Studies 2 and 3 in the present series of investigations were unable to demonstrate any beneficial effects of confirmation on either program errors or post-program retention. Other effects which may generally be assigned to confirmation were also conspicuously absent. The program completion time data appeared to preclude any indication of the possibility that subjects who did

not receive knowledge of results developed the compensatory strategy of spending more time in studying program frames. Moreover, no evidence was observed relating the presence vs. the absence of confirmation with subject loss during program administration.

As for the influence of response mode and number of required responses per frame, the only reliable main effects noted were on the combined unit tests and the post-program comprehensive test. Both treatments, requiring overt responses and increasing the number of responses per frame, enhanced retention scores for medical term responses, but did not generally affect scores based upon responses concerned with the definitions of medical terms. The differential effect occasioned by the manipulation of these variables is another finding that cannot be readily accounted for by operant conditioning principles.

The present findings, as well as those contained in the first two studies of this report, indicate that a strict adherence to the operant conditioning paradigm leads to an oversimplified conception of the associative learning processes that occur within the framework of programmed learning. It is now generally recognized by verbal learning theorists that the process of developing an associative connection, rather than it being simply one, involves a number of concurrent, but separate processes (Underwood and Schulz, 1960; McGuire, 1961). This perspective implies that the formation of an association between a learner's response and the eliciting material could be largely independent of the processes involved in learning the stimulus material or the response terms per se. Findings from studies conducted to determine the extent of stimulus item recall after paired-associate learning have a direct bearing on this matter. A procedure conventionally designated as R-S recall is used in such investigations. After a paired-associate list is learned, subjects are presented with the individual response terms (R's) and are requested to recall the stimulus terms (S's) with which the responses were paired during acquisition training. These studies have demonstrated that while subjects can learn to make the proper associations and master the response terms, as evidenced by the attainment of some learning criterion, they are still often unable to accurately reproduce the stimulus terms that were used to elicit the responses. Feldman and Underwood (1957), for example, found that with nonsense syllables as the stimulus components in paired-associate learning, only 50% of the stimulus terms were correctly recalled and accurately reproduced. Of considerable relevance to the present findings is the demonstration that stimulus recall is dependent upon the meaningfulness of the stimulus materials used. Hunt (1959), using lists compiled by Noble (1952), found that R-S recall ranged from 99% with highly meaningful stimulus material to 54% with stimuli rated low in meaningfulness. The relationship between amount of stimulus recall and meaningfulness has also been reported by Jantz and Underwood (1958) and Cassem and Kausler (1962).

In his analyses of verbal learning experiments, Underwood (1963) found it desirable to distinguish between nominal and functional stimuli, and this distinction is especially relevant to the R-S recall findings. When R-S recall is impaired, subjects are unable to recall the nominal stimulus; specifically, they cannot provide an accurate letter-by-letter reconstruction of the stimulus item as it appeared during the instructional session. Subjects in verbal learning experiments apparently do not always have to attend to nominal stimulus items as integrated units as long as there is no need to use these items in any overt response. Instead, any portion or characteristic of a particular item can be employed as a cue in forming an associative connection. To the extent that the cue utilized allows a subject to differentiate among all of the items serving as stimuli in the learning situation, the fractional component or characteristic selected from the experimentally presented, or nominal stimulus can become an effective functional stimulus. Given such a state of affairs, then, associative learning could take place in the absence of nominal stimulus item recall. The dynamics involved in such a cue selection process during paired-associate learning have been extensively documented (Weiss and Margolius, 1954; Hill and Wickens, 1962; Underwood, Ham and Ekstrand, 1962; Cohen and Musgrave, 1964; Houston, 1964; Jenkins and Baily, 1964; James and Greeno, 1967; Postman and Greenbloom, 1967).

It is apparent that words with little or no meaning can effectively elicit verbal or written responses even though the learner is unable to provide an accurate transcription of the words when requested to do so. If the frame content in self-instructional programs contains complex or technical terms that are initially unfamiliar to the learner, cue selection may occur if the terms are not required as overt program responses. Rather than attending to a technical term as a unified word, an individual can selectively focus upon certain combinations of letters to the exclusion of others. Any easily discernable characteristic of a term that would enable a learner to distinguish it from other terms in the program could provide the basis for cue selection. Under these conditions it would be expected that on later occasions individuals would be able to recognize the technical terms they had encountered in the program, but they would be unable to accurately reproduce them since they were not initially learned as integrated units. On the other hand, the meaningful words associated with the technical terms in the program (their definitions, for example) could be accurately reproduced when recalled since they would represent units well-integrated prior to their usage in the program. Results reported by Fry (1960) and Williams (1965) clearly support these contentions. In comparisons between constructed response and multiple-choice modes, these investigators found no differences when multiple-choice tests were used to measure retention of programmed subject matter. On retention tests requiring the written recall of technical terms, however, the constructed response groups displayed superior achievement. Moreover, in the investigation by Williams, no significant differences between response mode groups were observed on tests requiring either the recall or the recognition of the non-technical material learned in the program.

The retention test items in the present study required the learner not only to recall the association between a medical term and its definition, but to accurately reproduce the associative response as well. If medical terms were not properly integrated during the programmed learning session, their appearance as test responses would not correspond with the way they were spelled in the program. Depending upon the cue selection process used during learning, the test responses could range from slight misspelling to totally inaccurate reproductions. While the medical term scores on the combined unit tests in the present study showed the differences between overt and covert responding, as well as between overt responding and reading, to be significant regardless of spelling accuracy, a closer examination of the means for these differences showed a progressive diminution of the response mode effect as response reproduction accuracy was made less stringent. The effect of accuracy upon response mode comparisons was even more clearly demonstrated in the comprehensive test. The overt response mode was found to differ significantly from the covert and reading groups only when relatively high standards of spelling accuracy were demanded. When test responses below these standards were allowed, all response mode groups showed equivalent performances.

Definitions of medical terms do not require the training for integration that is necessary for the proper reproduction of medical terms. Words like "bladder" and "rupture", for example, are already in the college freshman's vocabulary, and if associated with the proper medical term, they can be accurately reproduced when recalled. Further, definitional phrases, unlike medical terms, can be integrated in various ways. The medical term "cystorrhexis", for example, requires learning a fixed sequence of letters for proper integration, but the range of acceptable definitions for this term includes (a) a ruptured bladder, (b) a bladder that is ruptured, (c) a rupture of the bladder, etc. The spelling of words like "hemmorrhage" and "inflammation", however, while not being as unfamiliar as most medical terms, would still benefit from procedures that result in response integration.

The results comparing response modes on test items requiring definitions of medical terms as responses were consistent with the position stated above. In sharp contrast to the widespread differences obtained with medical terms, test scores based on definition responses were not found to vary with response mode except in one instance. When the scoring procedure required accurate spelling of each word in the definitional phrase, the overt response mode subjects obtained higher scores on the combined unit tests.

Regardless of whether programmed learning follows a paired-associate or a serial learning paradigm, response integration training takes place when stimulus or response terms have to be overtly constructed during the learning process. When the terms are familiar and meaningful to begin with, or when they can be easily assimilated into transcribable units, overt response training is unnecessary for response integration. Thus, through writing, the cue selection process that occurs with covert responding and reading does not take

place, and technical terms that are originally a series of discreet letters or syllables become organized into recallable units.

From all indications it appears that the conflicting results in studies comparing overt responding with other response modes can be attributed to differences among studies in the degree of integration training required by both the program responses and the criterion test items. While response mode does not appear to be an important variable governing associative learning in self-instructional programs, it is a very definite factor in response learning. The major difficulty experienced by the subjects in the covert response and reading groups in the present study was not their inability to recall the association between medical terms and their definitions, but rather to accurately reproduce the medical terms themselves. This was evidenced by the failure of response mode to differentiate among the groups when definition test item responses were considered, and by the convergence of the mean scores among the response mode groups when increasingly less emphasis was placed upon spelling accuracy in scoring the medical term test item responses. Viewed in this light, the significantly longer times taken by the subjects in the covert response and reading groups in completing the retention tests would seem to be due, not to difficulties experienced in associative recall, but rather to difficulties in response reproduction.

Whatever the articulatory and corresponding mental processes in sub-vocal responding may be, it appears that they are no more effective than reading in promoting the recall of programmed material. The results obtained from the reading program groups closely paralled those of the covert response groups. In one instance, however, the Multiple Response Frame Program findings do raise some questions concerning the nature of covert responding. Increasing the number of responses required by the program frames produced higher scores on the combined unit tests and the comprehensive test, but only for the medical term test responses. Since the Multiple Response Frame Programs required 359 more responses concerned with the reproduction of medical terms than their single response counterparts, it would be consistent with the response integration interpretation to attribute the effectiveness of the experimental treatment to the increased frequency of response occurrence. However, it should be noted that no significant interaction between number of frame responses and response mode was obtained. The increase in medical term test response scores occurred for covert as well as overt responding. Obviously, the increased opportunity for covert responding was of some beneficial value, notwithstanding the fact that the increased covert response group scores did not significantly exceed those obtained by the reading program groups. The determination of the mechanism underlying the effect is not immediately apparent from the current series of investigations and requires further study.

Further research is also necessary to fully understand the implications of the delayed retention test data obtained in the present study. Contrary to the finding of Krumboltz and Weisman (1962), the scores on the delayed retention test were not found to be sensitive to

the effects of response mode. In fact, other than the significantly shorter test completion times exhibited by the subjects in the overt response groups, no other significant main or interactive effects were observed on this test. While the low delayed test scores may be said to be indicative of the overall insensitivity of the test, it is possible that the effects created by the manipulation of response mode and number of required responses per frame are too short-lived to influence performance on a test administered five weeks after program completion.

Finally, comment should be directed toward the role of confirmation in programmed learning. The fact that none of the different types of program errors analyzed in the present study were found to vary with the presence versus the absence of confirmation provides some suggestions concerning the utilization of feedback information in linear programs. Response integration learning could be expected to have been greatly facilitated if subjects went through the process of making point by point comparisons between their responses and the correct ones supplied by confirmation. There was no evidence that this occurred. The reproduction accuracy of the medical term program responses was not any more precise when confirmation was available that when it was deleted. Moreover, program errors in general, remained unaffected by the manipulation of confirmation.

Apparently, the confirmation item in self-instructional programs merely provides the learner with binary feedback information. It informs the learner whether his responses are "right" or "wrong". Ordinarily, it would be expected that association learning would be facilitated when this knowledge is provided. However, in self-instructional programs that are designed with continuous rather than discrete frame sequences, that deliberately feature much repetitive practice, that maintain a low error rate, and that provide for frequent review, the information supplied by confirmation is largely redundant and has no apparent instructional value.

References

Alter, M., and Silverman, R.E. The response in programmed instruction. Journal of Programmed Instruction, 1962, 1, 55-78.

Anderson, R.C. Educational Psychology. Annual Review of Psychology, 1967, 18, 129-164.

Barlow, J.A. The importance of overt response in programmed instruction as a function of Scholastic Aptitude Test scores. Paper presented to the American Psychological Association, Washington, D.C., September, 1967.

Barlow, J.A. The teacher and the teaching machine. AID, 1961, 1, 5-6.

Briggs, L.J. Learner variables and educational media. Review of Educational Research, 1968, 38, 160-176.

Cassem, N., and Kausler, D.H. Supplementary report: Effects of stimulus association value and exposure duration on R-S learning. Journal of Experimental Psychology, 1962, 64, 94.

Cohen, J.C., and Musgrave, B.S. Effect of meaningfulness on cue selection in verbal learning. Journal of Experimental Psychology, 1964, 68. 284-291.

Coulsen, E., and Silberman, F. Effects of three variables in a teaching machine. Journal of Educational Psychology, 1960, 51, 135-143.

Crist, R.L. Overt versus covert responding and retention by sixth-grade students. Journal of Educational Psychology, 1966, 57, 99-101.

Cummings, A., and Goldstein, L.S. The effect of overt and covert responding on two kinds of learning tasks. In DeCecco, J.P. (Ed.), Educational Technology. Holt, Rhinehart, and Winston, 1964, 231-241.

Dunn, O.J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.

Eigen, L.D., and Margulies, S. Response characteristics as a function of information level. Journal of Programmed Instruction, 1963, 2, 45-54.

Evans, J.L., Glaser, R., and Homme, L.E. An investigation of "teaching machine" variables using learning programs in symbolic logic. United States Office of Education, 1960.

Evans, J.L., Glaser, R., and Homme, L.E. An investigation of variations in the properties of verbal learning sequences of the "teaching machine" type. In Lumsdaine, A.A., and Glaser, R. (Eds.), Teaching Machines and Programmed Learning: A Source Book. Washington, D.C.: National Educational Association, 1960, 446-451. (a)

Estes, W.K. Learning. In Harris, C.W. (Ed.), Encyclopedia of Educational Research, 3rd ed., New York: The Macmillan Co., 1960, 752-770.

Feldhusen, J.F., and Birt, A. A study of nine methods of presentation of programmed learning material. The Journal of Educational Research, 1962, 55, 461-466.

Feldman, M.E. Learning by programmed and test format at three levels of difficulty. Journal of Educational Psychology, 1965, 56 133-139.

Feldman, S.M., and Underwood, B.J. Stimulus recall following paired-associate learning. Journal of Experimental Psychology, 1957, 53, 11-15.

Fiks, A.I. Some treatment and population variables in programmed instruction. Journal of Educational Psychology, 1964, 55, 152-158.

Fry, E.B. A study of teaching machine response modes. In Lumsdaine, A.A., and Glaser, R. (Eds.), Teachine Machines and Programmed Learning: A Source Book. Washington, D.C.: National Education Association, 1960, 469-474.

Gagne, R.M., and Bassler, O. Study of retention of some topics of elementary nonmetric geometry. Journal of Educational Psychology, 1963, 54, 123-131.

Gagne, R.M., and Smith, E.C., Jr. A study of the effects of verbalization on problem solving. Journal of Experimental Psychology, 1962, 62, 12-18.

Goldbeck, R.A., and Campbell, V.N. The effects of response mode and response difficulty on programmed learning. Journal of Educational Psychology, 1962, 53, 110-118.

Grace, H.A., and Cantor, J.M. Differential effectiveness of programmed presentation. Journal of Programmed Instruction, 1966, 3, 16-19.

Guthrie, E.R. The Psychology of Learning, rev. ed. New York: Harper, 1952.

Hall, M.E. Remote associative tendencies in serial learning. Journal of Experimental Psychology, 1928, 11, 65-76.

Hamilton, N.R., and Porteus, B.D. Instruction from small step vs. large step programs with immediate vs. delayed review procedures for students of different ability levels. Unpublished report, American Institute for Research, Palo Alto, 1965.

Hartman, T.F., Morrison, B.A., and Carlson, M.E. Active responding in programmed learning materials. Journal of Applied Psychology, 1963, 343-347.

Hill, F.A., and Wickens, D.D. The effect of stimulus compounding in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 563-566.

Holland, J.G. Design and use of a teaching machine program. Paper presented to the American Psychological Association, Chicago, September, 1960.

Holland, J.G. Research on programming variables. In Glaser, R. (Ed.), Teaching Machines and Programmed Learning II: Data and Directions. Washington, D.C.: National Education Association, 1965, 66-117.

Holland, J.G., and Kemp, F.D. A measure of programming in teaching-machine material. Journal of Educational Psychology, 1965, 56, 264-269.

Holland, J.G., and Skinner, B.F. The Analysis of Behavior. New York: McGraw-Hill, 1961.

Hough, J.B., and Revsin, B. Programmed instruction at the college level: A study of several factors influencing learning. Phi Delta Kappa, 1963, 44, 286-291.

Houston, J.P. S-R stimulus selection and strength of R-S association. Journal of Experimental Psychology, 1964, 68, 563-566.

Hughes, J.L. Effect of changes in program text format and reduction in classroom time on the achievement and attitude of industrial trainees. Journal of Programmed Instruction, 1962, 1, 43-54.

Hunt, R.G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. Journal of Experimental Psychology, 1959, 57, 262-267.

Jacobs, J.N., Yeager, H., and Tilford, E.J. An evaluation of programmed instruction for the teaching of facts and concepts. Journal of Programmed Instruction, 1966, 3, 29-38.

Jacobs, P.I., and Kulkarni, S. A test of some assumptions underlying programmed instruction. Psychological Reports, 1966, 18, 103-110.

James, C.T., and Greeno, J.G. Stimulus selection at different stages of paired-associate learning. Journal of Experimental Psychology, 1967, 74, 75-83.

Jantz, E.M., and Underwood, B.J. R-S learning as a function of meaningfulness and degree of S-R learning. Journal of Experimental Psychology, 1958, 56, 174-179.

Jenkins, J.J., and Bailey, V.B. Cue selection and mediated transfer in paired-associate learning. Journal of Experimental Psychology, 1964, 67, 101-102.

Kemp, F.D., and Holland, J.G. Blackout ratio and overt responses in programmed instruction: Resolution of disparate results. Journal of Educational Psychology, 1966, 57, 109-114.

Krumboltz, J.D., and Bonawitz, B. The effect of receiving the confirming response in context in programmed material. The Journal of Educational Research, 1962, 55, 472-475.

Krumboltz, J.D., and Weisman, R.G. The effect of intermittent confirmation in programmed instruction. Journal of Educational Psychology, 1962, 53, 250-253.

Krumboltz, J.D., and Weisman, R.G. The effect of overt vs. covert responding to programmed instruction on immediate and dealyed retention. Journal of Educational Psychology, 1962, 53, 89-92.

Lambert, P., Miller, D.M., and Wiley, D.E. Experimental folklore and experimentation: The study of programmed learning in the Wauwatosa Public Schools. The Journal of Educational Research, 1962, 55, 485-494.

Lepley, W.M. Serial reactions considered as conditioned reactions. Psychological Monographs, 1934, 46, No. 25.

Lublin, S.C. Reinforcement schedules, scholastic aptitude, autonomy need, and achievement in a programmed course. Journal of Educational Psychology, 1965, 56, 295-302.

Lumsdaine, A.A., and May, M.A. Mass communications and educational media. Annual Review of Psychology, 1965, 16, 475-513.

May, M.A. The role of student response in learning from new educational media. Office of Education, Bureau of Research, Department of Health Education and Welfare, Washington, D.C., August, 1966.

McGeoch, J.A., and Irion, A.L. The Psychology of Human Learning. Longmans, Green and Co., Inc., 1952.

McGuire, W.J. A multiprocess model for paired-associate learning. Journal of Experimental Psychology, 1961, 62, 335-347.

Meyer, S.R. Report on the initial test of a junior high school vocabulary program. In Lumsdaine, A.A., and Glaser, R. (Eds.), Teaching Machines and Programmed Learning: A Source Book. Washington, D.C.: National Education Association, 1960, 229-246.

Moore, J.W., and Smith, W.I. Knowledge of results in self-teaching spelling. Psychological Reports, 1961, 9, 717-726.

Noble, C.E. An analysis of meaning. Psychological Review, 1952, 59, 421-430.

Postman, L., and Greenbloom, R. Conditions of cue selection in the acquisition of paired-associate lists. Journal of Experimental Psychology, 1967, 73, 91-100.

Reid, R.L., and Taylor, J. Programmed learning with and without responding. Occupational Psychology, 1965, 39, 145-149.

Ripple, R.E. Comparison of the effectiveness of a programmed text with three other methods of presentation. Psychological Reports. 1963, 12, 227-237.

Roe, A. Automated teaching methods using linear programs. Journal of Applied Psychology, 1962, 40, 198-201.

Rosenstock, E.H., Moore, J.W., and Smith, W.I. Effects of several schedules of knowledge of results on mathematics achievement. Psychological Reports, 1965, 17, 535-541.

Shay, C.B. Relationship of intelligence to step size on a teaching machine program. Journal of Educational Psychology, 1961, 52, 98-103.

Silberman, H.F., Melaragno, R.J. and Coulson, J.E. Confirmation and prompting with connected discourse material. Psychological Reports, 1961, 9, 235-238.

Skinner, B.F. Teaching machines. Science, 1958, 128, 969-977.

Skinner, B.F. The programming of verbal knowledge. In Galanter, E.H. (Ed.), Automatic Teaching: The State of the Art. New York: John Wiley and Sons, 1959, 63-68.

Smith, G.L., and Davis, P.E. Medical Terminology: A Programmed Text. New York: John Wiley and Sons, 1963.

Smith, W.I., and Moore, J.W. Size of step and achievement in programmed spelling. Psychological Reports, 1962, 10, 287-294.

Stewart, D.J., and Chown, S.M. A comparison of the effects of a continuous and a linear programmed text on adult pupils. Occupational Psychology, 1965, 39, 135-144.

Stolurow, L.M., and Walker, C.C. A comparison of overt and covert response in programmed learning. The Journal of Educational Research, 1962, 55, 421-432.

Tobias, S., and Weiner, M. Effect of response mode on immediate and delayed recall from programmed materials. Journal of Programmed Instruction, 1963, 2, 9-18.

Underwood, B.J. Stimulus selection in verbal learning. In Cofer, C. N., and Musgrave, B.S. (Eds.), Verbal Learning and Verbal Behavior, New York: McGraw-Hill, 1963, 33-48.

Underwood, B.J., Ham, M., and Ekstrand, B. Cue selection in paired-associate learning. Journal of Experimental Psychology, 1962, 64, 405-409.

Underwood, B.J., and Schulz, R.W. Meaningfulness and Verbal Learning. Philadelphia: Lippencott, 1960.

Warren, A.D. To program or not to program. Journal of Programmed Instruction, 1966, 3, 41-44.

Weiss, W., and Margolius, G. The effect of context stimuli on learning and retention. Journal of Experimental Psychology, 1954, 48, 318-322.

Williams, J.P. Combining response modes in programmed instruction. Journal of Educational Psychology, 1966, 57, 215-219.

Williams, J.P. Comparison of several response modes in a review program. Journal of Educational Psychology, 1963, 54, 253-260.

Williams, J.P. Effectiveness of constructed-response and multiple-choice programming modes as a function of test mode. Journal of Educational Psychology, 1965, 56, 111-117.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Wittrock, N.C. Response mode in the programming of kinetic molecular theory concepts. Journal of Educational Psychology, 1963, 54, 89-93.

Yarmey, A.D. Overt and covert responding in programmed learning. The Journal of Educational Research, 1964, 7, 27-33.